Disjoint v Independent

It is very important to understand the difference between disjoint and independent events. Expect at least one question that involves you understanding the difference.

- **Disjoint:** Mutually exclusive. Two events cannot happen at same time.
  - $P(A \cap B) = 0$
- **Independent:** The outcome of one event does not influence the outcome of the other.
  - $P(A|B) = P(A)$
  - $P(A \cap B) = P(A) \cdot P(B)$
  - Any time you see independence given in a problem, write down $P(A \cap B) = P(A) \cdot P(B)$, it will probably be used in the problem.

8. The probability of event A is 0.3. The probability of event B is 0.6. If events A and B are disjoint, then:
   A. P(A or B) = 0.9 B. P(A and B) = 0.18
   C. P(A or B) = 0.72 D. P(A or B) = 0

$P(A \cap B) = 0$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= .3 + .6 - 0$
$= .9$

9. If P(A) = 0.30 and P(B) = 0.25, what is P(A or B) if A and B are independent?
   A. 0.075 B. 0.55 C. 0.475 D. 0 E. None of these

$P(A \text{ OR } B) = P(A) + P(B) - P(A \cap B)$
$= .3 + .25 - .075 = .475$

$P(A \cap B) = P(A) \cdot P(B)$
$= (.3)(.25) = .075$

10. Two events, X and Y, are independent, such that P(X) = 0.41 and P(Y) = 0.52. What is the value of $P(X \cup Y)$?
    A. 0.2132 B. 0.9300 C. 0.1100 D. 0.7168

same as #9.

Given Formulae

- $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Should be used for any "or" question.
  - Can also be used to solve for "and" with a little algebra:;
    $P(A \cup B) = P(A) + P(B) - \mathbf{P(A \cap B)}$

- $P(A \text{ and } B) = P(A \cap B) = P(A|B) * P(B)$
  - Only use if you have a conditional! Otherwise, try using the formula above.
  - If you have P(B|A), don't fret: $P(A \cap B) = P(B \cap A)$,
    so $P(A \cap B) = P(B|A) * P(A)$

$P(D) = .39$ $P(G) = .21$ 1-none P(none) = .53

11. 39% of new model cars have a built in DVD player, 21% have a GPS navigation system, and 47% have at least one of those features. What is the probability that a new model car has both these features?
    A. 0.13 B. 0.08 C. 0.26 D. 0.0819 E. 0.6

$P(D \text{ OR } G) = .47$

$P(D \text{ OR } G) = P(D) + P(G) - P(D \text{ and } G)$
$.47 = .39 + .21 - x$ $x = .13$

12. Advertisements for statistics texts suggest that 79% come with a CD with statistical tools, 35% have online help, and 18% have both aids. Find the probability that a randomly selected statistics text has statistical tools or online help.
    A. 1.14 B. 0.18 C. 0.96 D. 0.28 E. 0.61

S and H S H S ∪ H

$P(S \text{ OR } H) = P(S) + P(H) - P(S \text{ and } H)$

Tree Diagrams

If you're given two conditional probabilities and one marginal probability, it's likely a tree diagram will help you solve the question! If it is stated on the test that a tree diagram might help, DRAW A TREE DIAGRAM!

- **Memorize** the probability notation and construction of the tree. The items in blue below will be given (or can be calculated by subtracting a given probability from 1).
- You can use the probabilities in green to fill in a contingency table and then answer ANY question!

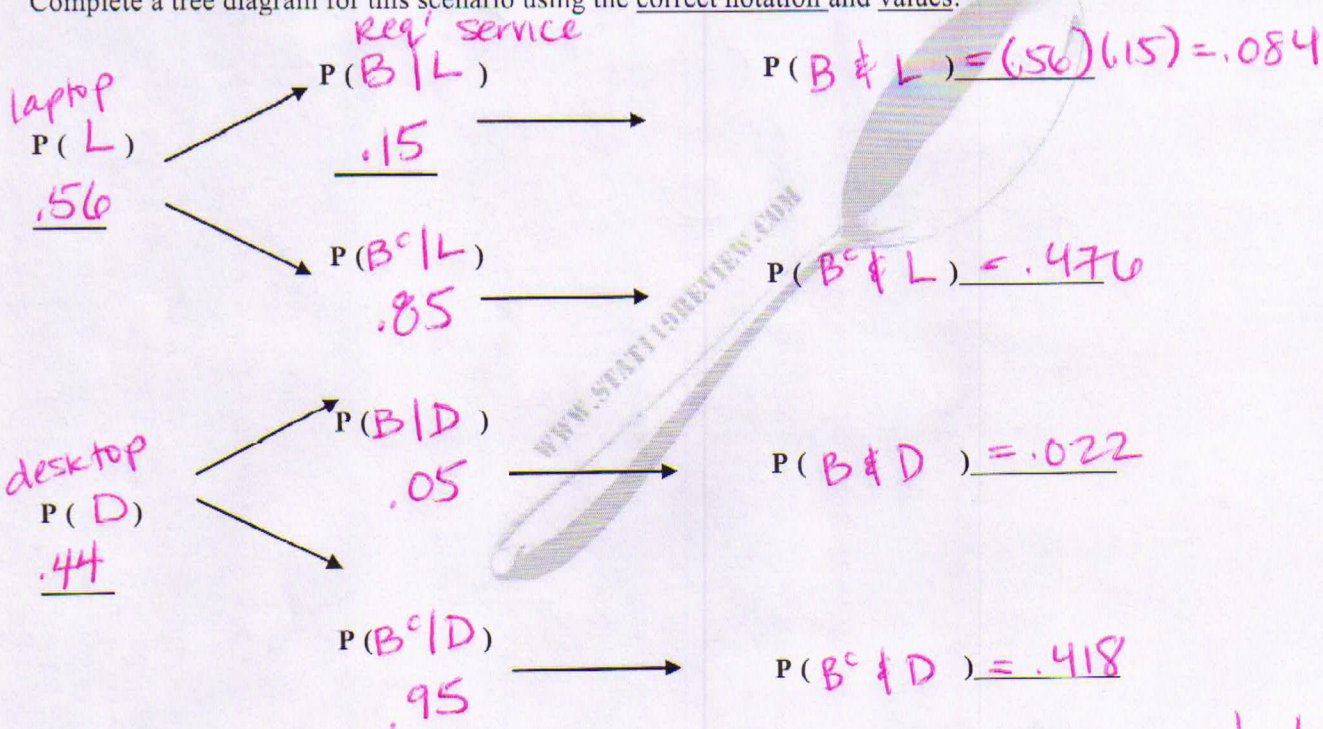
Hints and Tips:

- The most common questions asked in tree diagrams are for P(B) and P(A|B) or a similar type of probability.
- For P(B), just add the two joints in the green section that have a B in them.
- For P(A|B), the given item is almost always the thing you calculated first, so you can reuse your work!

Use the following information for questions 12-13

Suppose a campus computer store sells laptop and desktop computers. The probability a computer sold is a laptop computer is 0.56 (event A). In the first year, 15% of laptop computers require service, while 5% of desktops require service. Let B be the event that a computer requires service.

Complete a tree diagram for this scenario using the correct notation and values:



	L	D	
B	.084	.022	.106
B^c	.476	.418	.894
	.56	.44	1

13. What is the probability a computer required service and was a laptop?

A. 0.028 (B. 0.084) C. 0.28 D. 0.2 E. None of these

14. What is the probability that a computer will require service?

P(B) = .106

15. Given that a computer required service, what is the probability that it was a laptop?

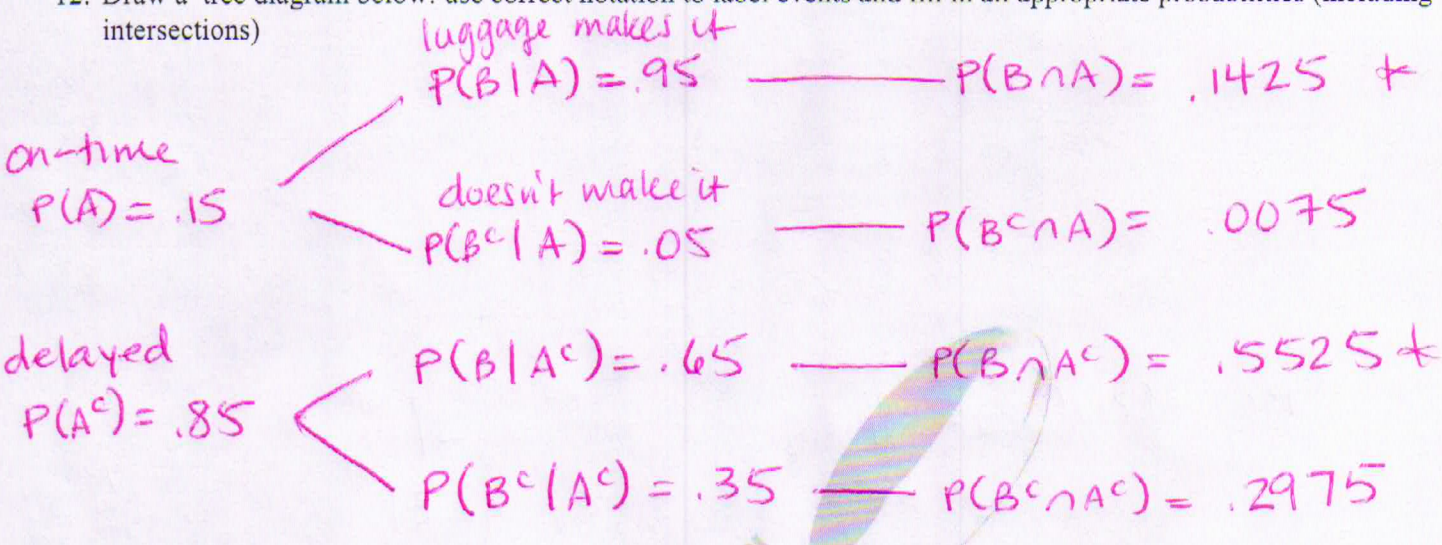
(A. 0.792) B. 0.084 C. 0.298 D. 0.42 E. None of these

Use the following to answer questions 12-14

Leah is flying from Boston to Denver with a connection in Chicago. The probability her first flight leaves on time is 0.15. If the flight is on time, the probability that her luggage will make the connecting flight in Chicago is 0.95, but if the first flight is delayed, the probability that the luggage will make it is only 0.65. Let A be the event that the first flight leaves on time, let B be the event that the luggage will make the connecting flight.

$P(A) = .15$ $P(B|A) = .95$ $P(B|A^c) = .65$

12. Draw a tree diagram below: use correct notation to label events and fill in all appropriate probabilities (including intersections)



13. What is the probability that her luggage arrives in Denver with her?

$P(B) = .695$ $.1425 + .5525$

14. If her luggage does not arrive in Denver with her, what is the probability her first flight did not leave on time?

$$P(A^c|B^c) = \frac{P(A^c \cap B^c)}{P(B^c)} = \frac{.2975}{.305} = .9754$$

Sampling without Replacement

Recognizing Sampling without Replacement

- The simplest question will say, "without replacement."
- A more difficult question will leave these keywords out. Instead, you'll need to notice a few key things.
  - The number of total items will be given, and likely fairly small < 30.
  - We will want to come away with two items, not just recording a characteristic of those two items. Selecting a committee and purchasing items are good examples of this.

Hints and Tips for Sampling without Replacement:

- Most common thing asked for is the probability of getting EXACTLY 1 of a type of item, do not forget that there are two ways to pick this if we choose 2 items!!
- Remember, in sampling without replacement, you need to determine how the first pick effects the probability in the second pick.

16. In a group of eleven adults, there are three who are allergic to peanuts. If we select two adults at random from the group (without replacement), what is the probability we select exactly one who is allergic to peanuts?

A. 0.3967 B. 0.2182 C. 0.2727 D. 0.4364 E. None of these

$$2 \cdot \frac{3}{11} \cdot \frac{8}{10} = .4364$$

17. A box contains 16 soccer jerseys: 4 medium, 7 large and 5 extra large. Two jerseys are drawn from the box, without replacement. What is the probability exactly 1 medium jersey is selected?

A. 0.375 B. 0.20 C. 0.05 D. 0.40

$2 \cdot \frac{4}{16} \cdot \frac{12}{15} = .40$ (M, M^c)

Use the following to answer 18 and 19:

Suppose a fun-size bag of skittles has 5 red, 7 green, 6 purple and 8 yellow (a total of 26 skittles). You select two skittles from the bag, one at a time, without replacement.

18. What is the probability that both of the skittles are green?

A. 0.0646 B. 0.0725 C. 0.2692 D. 0.5092

$1 \cdot \frac{7}{26} \cdot \frac{6}{25} = .0646$ (G, G)

19. The probability that exactly one of the two skittles is red is:

A. 0.037 B. 0.162 C. 0.323 D. 0.311 E. 0.155

$2 \frac{5}{26} \cdot \frac{21}{25} = .323$ (R, R^c)

20. Suppose you win a contest and as a prize are allowed to draw, while blindfolded, two bills from a container. Inside the container are seven \$1 bills, four \$10 bills, and eight \$20 bills. What is the probability of drawing at least 1 \$20 dollar bill?

A. 0.515 B. 0.257 C. 0.421 D. 0.678

1 − none $= 1 - \frac{11}{19} \cdot \frac{10}{18} = .678$ (not 20, not 20)

Sampling with Replacement

- The simplest question will say "independence" somewhere in the question.
- A more difficult question will say something about picking a small number of people using a simple random sample (or the word normal), but no total number of people is mentioned.

Hints and Tips for Sampling with Replacement:

- One of the most common things asked for is P(at least one), remember, you should calculated this using P(at least one) = 1 – P(none)
- Remember also, P(none) is P(not whatever) raised to the power of however many things we selected. A lot of students mistakenly use the given probability and forget to subtract it from 1 first!

17. Eighty-four percent of households in the United States own a computer. A random sample of four households is selected. What is the probability that none of the households own a computer?

A. 0.4979 B. 0.0026 C. 0.0007 D. 0.0138 E. None of these

$(.16)(.16)(.16)(.16) = .16^4$ (C^c, C^c, C^c, C^c)

18. Assume we know that the probability that a college student has a Twitter account is 0.18. Three students are selected at random (therefore we can assume independence). What is the probability that at least one of these has a Twitter account?

A. 0.5514 B. 0.4486 C. 0.0058 D. 0.9942

1 − none

$1 - (.82)(.82)(.82) = 1 - .82^3$ (TC, TC, TC)

Probability Formulas:

$P(A^c) = 1 - P(A)$

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \text{ and } B) = P(A \cap B) = P(A \mid B) \bullet P(B)$

$P(A \mid B) = \frac{P(A \cap B)}{P(B}$

Probability Distributions

A probability distribution is just a way to organize and show all possible outcomes and the chance of them occurring.

Sampling without replacement is the most common type of discrete probability distribution tested.

Creating a Probability Distribution

Step 1: Write a table with X on top and P(X) on the bottom.
Step 2: Fill in the remaining the possible outcomes X.
Step 3: Determine the probability of each outcome X and fill them into the table.
(Pay careful attention to whether you are dealing with sampling with or without replacement!)

Ex: A junk drawer in your house contains 15 old batteries, 5 of which are totally dead. You need two batteries for your remote control, so you select two batteries from the drawer.
Give a table with the probability distribution for X = the number of good batteries you get out of the drawer.

$P(X=0) = \frac{5}{15} \cdot \frac{4}{14}$ (D D)

$P(X=2) = \frac{10}{15} \cdot \frac{9}{14}$ (G G)

X	0	1	2
P(X)	.0952	.4762	.4286

$P(X=1) = 2 \cdot \frac{10}{15} \cdot \frac{5}{14}$ (G D)

Calculating the Mean and Standard Deviation for a Discrete Probability Distribution

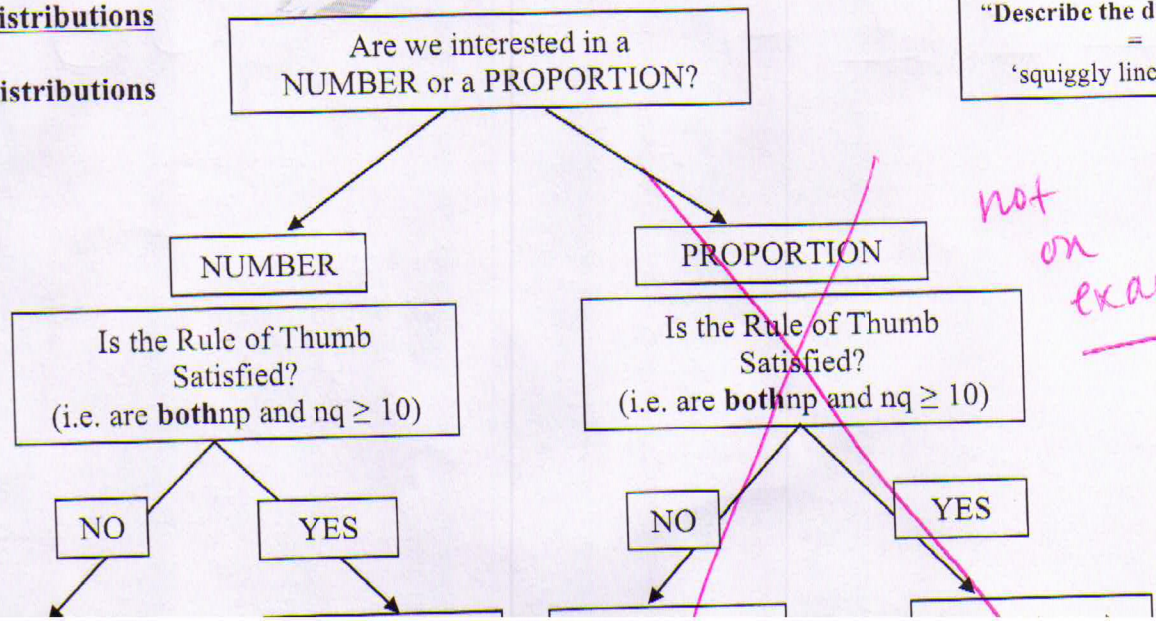
Step1: Enter all the X values in L1 and all the probabilities (as decimals) in L2.
Step 2: 1-VAR-STATS L1, L2 ** full details under calculator shortcuts on stat119review.com

Ex: Calculate the mean and standard deviation for the probability distribution we created above.

Sampling Distributions

Sampling Distributions

"Describe the distribution" = 'squiggly line' notation



not on exam 2

n=20 → p=1/6 be BINOM

Ex: Identifying Sampling Distributions

1. Consider rolling a die 20 times and recording the number of 2's rolled. What is sampling distribution of X, the number of 2's rolled?

X ~ B(20, 1/6)

> Remember: "**proportion**" or "**number**"

2. An airline, knowing that about 5% of passengers fail to show up for flights, overbooks (sells more tickets than there are seats). Suppose the airline sells 275 seats for a flight. What is the sampling distribution of the number of passengers that will fail to show up?

X ~ AN(13.75, 3.6142)

μ = np = 275(.05) = 13.75

σ = √npq = √275(.05)(.95)

> For "number" questions, check the **Rule of Thumb**!

3. The Harvard College Alcohol Study finds that 67% of college students support efforts to "crack down on underage drinking." The administration of a college surveys 100 students and finds that 62% support a crackdown on underage drinking. Describe the distribution of the sample proportion of students who support a crackdown on underage drinking.

IGNORE FOR EXAM 2

> Be prepared for a question about assumptions or requirements!

Requirements & Assumptions of the Distributions

Properties of the Binomial Distribution

1. There are a fixed number (n) trials or observations.
2. The trials or observations are independent.
3. There are only two outcomes: Outcome of interest (success) and its complement (failure).
4. The probability of success is the same for each trial.

Properties of the Approximate Normal Distribution for a NUMBER

1. The above properties for a binomial must be satisfied.
2. The Rule of Thumb must be satisfied (np and nq ≥ 10).
3. Sample must come from SRS.

Properties of the Approximate Normal Distribution for a PROPORTION

1. Sampled values must be independent.
2. The Rule of Thumb must be satisfied (np and nq ≥ 10).
3. Sample must come from SRS.
4. Sample must be < 10% of population.

> Same for inference except we don't know p, so we use p-hat in our Rule of Thumb!

Ex: There are approximately 1200 students enrolled in Stat119. To estimate the proportion of Stat119 students who voted in the election, a simple random sample of 200 students was taken from all students enrolled and it was found that 90 students voted. Are the necessary conditions met to create a confidence interval?

A. Yes, all of the requirements have been met.
B. No, nq̂ is less than 10.
C. No, n is greater than 10% of the population.
D. No, the sample is not a simple random sample.

IGNORE

Ex: All of the following are requirements for using the normal approximation except:

A. Data must be collected using a simple random sample. ✓
B. The sample size must be at least 10% of the population size. not really on exam 2 ..
C. np and n(1-p) must be greater than or equal to 10. ✓
D. The sampled values must be independent from one another. ✓

Properties of a Normal or Approximately Normal Distribution

There normally aren't many questions about the properties of the normal distribution, but you should know that its values follow a*symmetricbell-curve*.

> Standardizing changes the mean to 0 and the standard deviation to 1. The shape of the distribution is not changed.

Determining Question Types

Before starting any question, you should determine what type of question it is. To help, here is a list of types of questions on Exam 2, not including those for inferential statistics (confidence intervals, hypothesis testing and sample size calculations) For greater detail, see the 'Sampling Distribution' handout on stat119review.com:

Normal Questions – any question that has the word "normal" in the question stem.
- **Direct** – given a value, asked for a probability or a percent
- **Inverse** – given a probability or a percent, asked for a value

Number / Proportion Questions – you'll be given n (usually a sample size) and p (as a decimal or percent, possibly even as a ratio.)
- **Number** – find a probability that less than 30 people attended a concert
  - **Binom** – if np OR nq< 10
  - **AN** – if np AND nq ≥10
- **Proportion** – find the probability that more than 52% of people use liquid handsoap

Normal Distribution

If the word "normal" is anywhere in the question, you'll be doing one of these types of questions!

A good first step is determining if it's a direct or inverse question!

Z-Scores / Standard Normal

A Z-score gives us the number of standard deviations away from the mean a value is, as well as if that value is below or above the mean. You can find the Z-score for a value with the given formula:

If you "standardized" your last test score and found you had a Z-score of 2, that would mean that yourscore is 2 standard deviations above the mean. A Z-score of -1 would mean that your score was 1 standarddeviation below the mean.

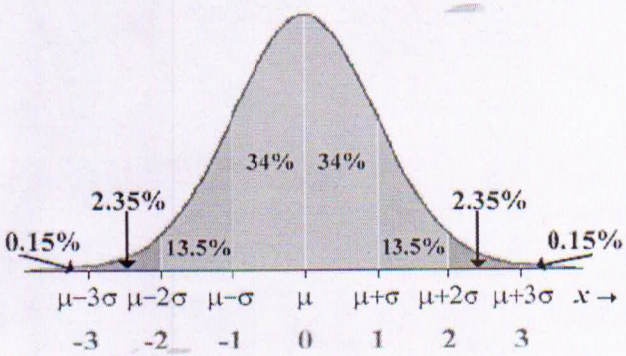
Two Types of Questions

Unusual: Scores further away from zero are more unusual, regardless of sign.
Better: Be careful about direction if asking who performed "better" – for a test, a higher Z-score would be better since scoring higher on the test is better; however, for a race, a lower Z-score is better since a smaller timeis actually a better performance.

The Empirical Rule

The Empirical Rule provides an estimate of the percent of data falling between certain values on a normal curve. Do NOT use the Empirical Rule unless the question expressly states to solve using the Empirical Rule.

Memorize the FOUR values corresponding to the amounts inside each section of half the curve.



Step 1: Draw your curve. Start with a line in the middle for your mean and then three additional lines to either side. Add standard deviations to the mean to get the values to the right. Subtract to get the values to the left. Then fill in the FOUR insidepercentages.

Step 2: Add up the percentages that correspond to portion of the curve you're being asked about.

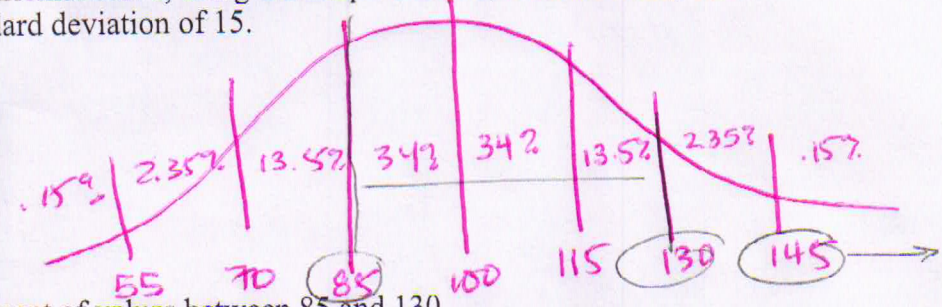
Hints and Tips:

- If you really hate the Empirical Rule (I do!), then you could always solve these problems as if they were a regular normal problem. Your answer will be just slightly off from the estimate provided by the Empirical Rule. (Only works on multiple choice.)
- If you are looking for values that aren"t the ones you"ve written at the bottom of the curve, you"ve done something wrong. Either you"re using the Empirical Rule when the question didn"t say to OR you made a simple math error when constructing your curve.
- For questions that give you a percentage and ask what the value is, first determine if they are talking about a value on the right side or the left side, then tackle it in the forward direction using the three values on that side of the curve.
- You may still be asked a question about the other three values: 68-95-99.7%, but it would only be in a multiple choice or true/false question.

Ex: True or false. Approximately 95% of the data in a normal distribution falls within 2 standard deviationsof the mean.

TRUE

Ex: Draw out the normal curve, using the Empirical rule, for a normally distributed variable with a mean of 100 and a standard deviation of 15.



Now, find the percent of values between 85 and 130.

$34+34+13.5=$ 81.5%

And the percent of values greater than 145.

.15%

Find value such that 16% are higher.

115

Direct Calculations

This is when we're given a value and asked to find a probability or percent above or below that value (or between two values). Normally this is as easy as looking at the question, did they ask you: "what is theprobability…" or "what percent"?

These problems boil down to the following sequence:

$$X \rightarrow Z \rightarrow \%$$

Step 0: Draw your curve! This stops you from choosing the opposite area. There will be a keyword letting you know which side you're interested in.
Step 1: $X \rightarrow Z$: Convert the given value to a Z-score using the **direct** formula in the formula sheet.
Step 2: $Z \rightarrow$ %: Convert that Z into a probability (area under the curve) by looking it up in the Z-table.Remembering that your table gives out the left side probability.

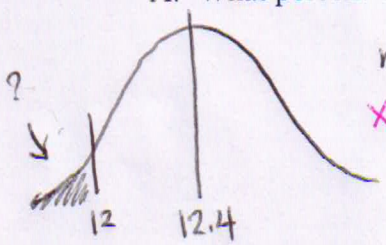
Calculator: normalcdf(lower, upper, mean, sd)

Ex: A pasta manufacturer has a machine that fills the boxes. The boxes are labeled "12 ounces" so the company wants to have that much pasta in each box. To prevent underweight boxes the manufacturer sets the mean fill higher than 12 ounces. Suppose the amount of pasta in the boxes follows a normal distribution with a mean fill of 12.4 ounces and a standard deviation of 0.24 ounces.

$\mu = 12.4$ $\sigma = 0.24$

Draw your curve. It will save you from wrong answer traps on multiple choice questions!

A. What percent of boxes weigh less than 12 ounces (i.e. are underweight)?

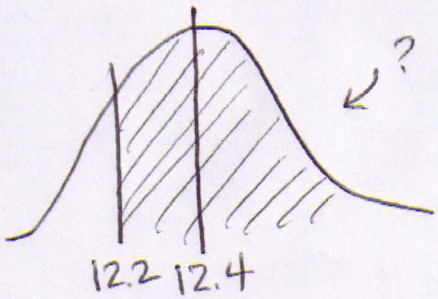


normalcdf(-999, 12, 12.4, 0.24) = .0478

$X \rightarrow Z$

$Z = \frac{12 - 12.4}{0.24} = -1.67$ $Z \rightarrow \%$ .0475

B. What percent of boxes weigh more than 12.2 ounces?



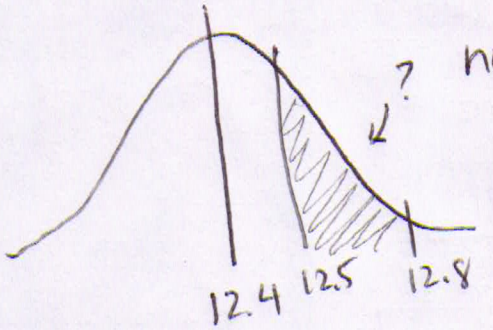
normalcdf(12.2, 9999, 12.4, 0.24) = .7977

$X \rightarrow Z$

$Z = \frac{12.2 - 12.4}{.24} = -.83$ $Z \rightarrow \%$ $1 - .2033 = .7967$

C. What is the probability that a randomly selected box weighs between 12.5 and 12.8 ounces?

On the between questions, you will get two Z-scores, look them both up and subtract the smaller one fromthe larger one.



normalcdf(12.5, 12.8, 12.4, 0.24) = .2907

$X \rightarrow Z$

$Z_2 = \frac{12.8 - 12.4}{.24} = 1.67 \rightarrow .9525$

$Z_1 = 12.5 - 12.4 = .42 \rightarrow .6628$ = .2897

Inverse Calculations

If at any point in the question they give you a percentage, percentile or quartile, you're doing an inverse question.These questions will usually also ask for a value.

Do not get these confused with the sample proportion questions! You need the word NORMAL before you even consider an inverse question.

These problems boil down to the following sequence:

$$\% \rightarrow Z \rightarrow X$$

Step 0: Draw your curve! This stops you from choosing the opposite value. There will be a keyword letting you know which side you're interested in.
Step 1: $\% \rightarrow Z$: Convert the given area to a Z-score using**inv**Norm in your calculator.Beware,invNorm takes the left side probability!
Step 2: $Z \rightarrow X$: Convert that Z into Xusing the **inverse** formula on your formula sheet .

Calculator: invNorm(left area/probability, mean, sd)

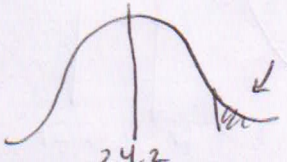
There will definitely be one of the inverse questions where you need to solve for the mean or standard deviation. This requires you to use the steps, there is no calculator shortcut!

Ex: The diameters of a certain airplane tire (for landing gear) are normally distributed with an average of 24.2 inches, and a standard deviation of 0.15 inch. $\mu = 24.2$ $\sigma = 0.15$

A. What is the probability that the tire's diameter is above 23.98 inches?

normalcdf(23.98, 9999, 24.2, 0.15) = .9288

B. How large would the diameter need to be for it to qualify as one of the 8% largest diameters? Right side!

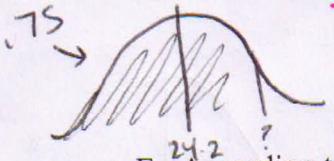


invNorm(.92, 24.2, 0.15) = 24.4108

$z \rightarrow z$: $z = 1.41$

$z \rightarrow x$: $x = 1.41(0.15) + 24.2 = 24.4115$

C. What is the third quartile for these tire diameters? 75th percentile



invNorm(.75, 24.2, 0.15) = 24.3012

Ex:According to the 2008 CIA World Factbook, the country with the world's longest life expectancy is Macau. Assume life expectancy is normally distributed and that 75% of people from Macau live to be at least 80 years old. If the standard deviation is 5 years, what is the mean life expectancy in Macau?

$x = 80$ $\sigma = 5$ $\mu = ?$

$\% \rightarrow Z$

$Z = \text{invNorm}(.25, 0, 1) = -.67$

$z \rightarrow \mu$

$x = z(\sigma) + \mu$

$80 = -.67(5) + \mu$

$\mu = 83.35$

Number Questions

If you're given n and p, then asked about a **number**, you're doing one of these types of questions.

Your first step needs to be to check the Rule of Thumb, to see if you're doing a BINOMIAL or APPROX NORMAL question. You can get answers using the wrong method, but they'll be wrong answers…

Binomial Questions

You'll see two probability questions on the binomial on Midterm 2 – probably an exact question and one inequality question.

Exact Questions – will ask the probability that a single number is our outcome.

Calculator: binompdf(n, p, k) calculates the probability of getting exactly k successes.

Note: If the binomial question is a free response, you are expected to write the formula from the formula sheet (withvalues input) even if you're using binompdf.

InequalityQuestions – will ask the probability that our outcome is part of a range of values

Step 1: Write out all possible outcomes: 0, 1, 2…. N
Step 2: Circle the outcomes that we're interested in.
Step 3: Use either binompdf to add up all outcomes we're interested in OR use binomcdf.

Calculator: binomcdf(n, p, k) adds up the probabilities from 0 up to and including k.

Note: It may be necessary to use the complement if it's easier to calculate. Be careful. Writing the list of outcomes will make sure you don't make a mistake.

Ex: Assume that 15% of people in the US are left-handed. If 10 people are selected at random (we can assume independence), find each of the following probabilities:
Probability that exactly 3 of the people selected are left-handed.

binompdf(10, .15, 3) = .1298

Probability that at least 3 people selected are left-handed.

0 1 2 (3 4 5 6 … 10)

1 − binomcdf(10, .15, 2) = 1 − .8202 = .1798

Probability that at most 8 people selected are left handed.

(0 1 2 3 4 5 6 7 8) 9 10

binomcdf(10, .15, 8) = .9999…

oops

Approximate Normal Questions

These are the same as direct normal question, but you need to first calculate the mean and standard deviation from the "number" section of your formula sheet.

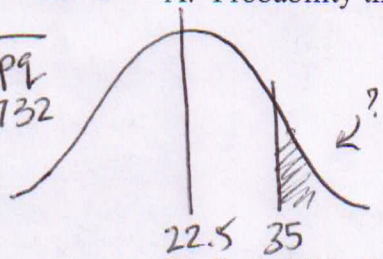
$n=150$ $p=.15$

$\mu=np=150(.15)$
$=22.5$

Ex. Assume that 15% of people in the US are left-handed. If 150 people are selected at random (we can assume independence), find each of the following probabilities:

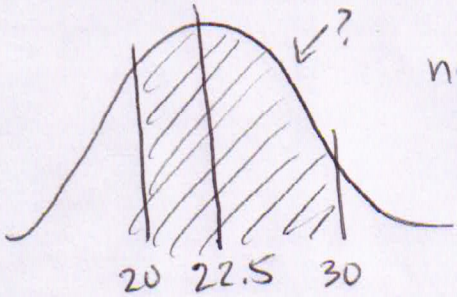
A. Probability that at least 35 of the people selected are left-handed.

$\sigma=\sqrt{npq}$
$=4.3732$



normalcdf(35, 99999, 22.5, 4.3732)
$=$ .0021

B. Probability that between 20 and 30 people selected are left-handed.



normalcdf(20, 30, 22.5, 4.3732)
$=$ .6731

Proportion Questions

If you're given n and p, then asked about a **proportion**, you're doing one of these types of questions.

These are the same as direct normal question, but you need to first calculate the mean and standard deviation from the "proportion" section of your formula sheet.

Ex. Assume that 25% of all business students at a large university invest in the stock market. A random sample of 110 business students is selected from this university. What is the probability that more than 34% of this sample invests in the stock market?

not on exam 2