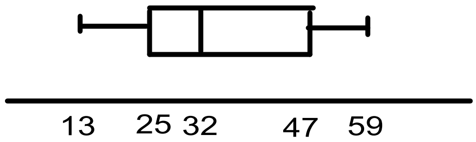
Statistics and Probability Review (Topic 6)

You must be able to work with frequency tables, diagrams, box and whisker plots.

Ex) Consider the following box and whisker plot. What does it tell us?

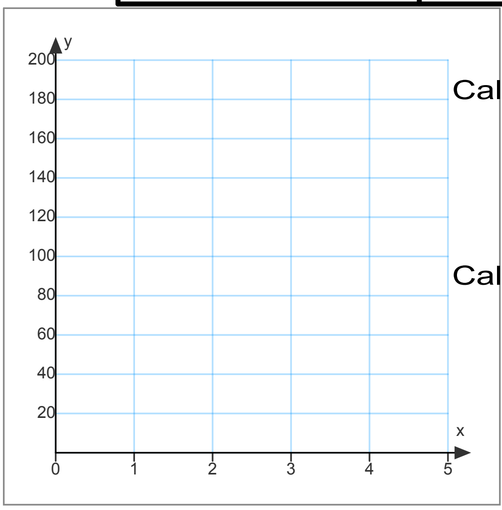


Ex) The lifespans of 200 pet goldfish are represented in the frequency table below.

Age (years)	0-1	1-2	2-3	3-4	4-5
Freq.	40	77	32	28	23

Complete the following cumulative frequency table and draw a cumulative frequency curve.

Age (years)	≤ 1	≤ 2	≤ 3	≤ 4	≤ 5
Cumulative freq.					



Calculate the expected lifespan of a goldfish.

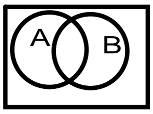
Calculate the standard deviation of goldfish lifespans.

Use your cumulative frequency distribution to estimate the median lifespan of a goldfish to one decimal place.

A goldfish is considered mature if it is more than two years old. A goldfish is selected at random. What is the probability that it is not mature?

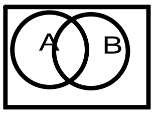
–

Probability Terms:



$A \cup B$ reads: "the union of sets A and B". It can be interpreted as A or B or both

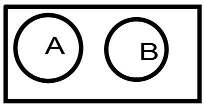
$A \cap B$ reads: "the intersection of sets A and B". It can be interpreted as both A and B



$P(A|B)$ reads: "the probability of event A, given that B has occurred".
Generally, $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Complementary events: if the event A is being late for school, then its complement, A', is not being late for school. Generally, $P(A) + P(A') = 1$

Mutually exclusive events: These are events which cannot occur simultaneously. Consider a deck of cards and let event A be drawing a Club and event B be drawing a red card. In a single draw, these events are mutually exclusive (a card cannot be both a Club and red). For mutually exclusive events, $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$



Not mutually exclusive events: These are events which can occur simultaneously. Consider a deck of cards and let event C be drawing a Heart and event D be drawing a Queen. In a single draw, these events are not mutually exclusive (a card can be both a Queen and a Heart). For not mutually exclusive events,

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *If you're not sure about exclusivity, use this formula for unions.*

Independent Events: These are pairs of events where the outcome of one event does not affect the probability of the other. Classify the following pairs of events as dependent or independent:

1) Tossing a coin and rolling a die.

2) Drawing two clubs from a deck without replacement.

For independent events, $P(A \cap B) = P(A)P(B)$

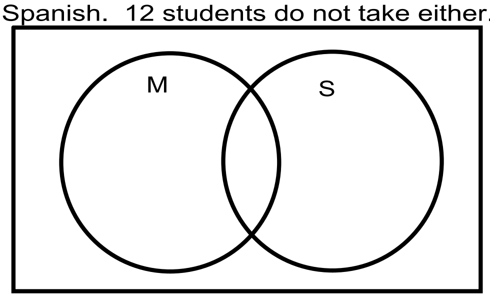
Conditional Probability involves dependent events and often "...given that..." statements. The guiding formula is called Bayes' Theorem and is given as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$ (explained above).

Probability

You should be able to work with Venn diagrams and tree diagrams.

Ex.) There are 82 students in grade 12 at a school. 64 students take math and 56 students take Spanish. 12 students do not take either.

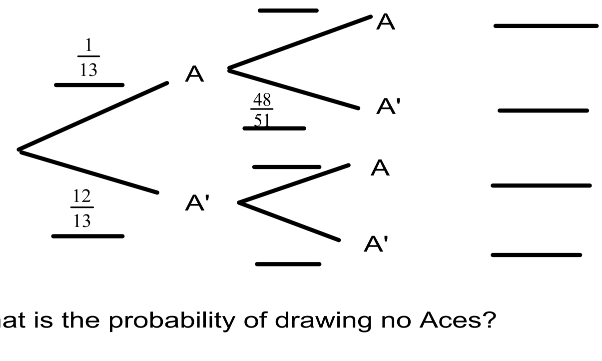


How many students take both math and Spanish?

Given that a student takes math, what is the probability he/she does not take Spanish?

Show that M and S are not independent.

Ex.) A standard deck of cards contains 52 cards and 4 Aces. A student selects a card from a standard deck and records whether or not it is an Ace. Without replacing the first card, he selects a second card and records whether it is an Ace. Complete the following tree diagram.



What is the probability of drawing no Aces?

What is the probability of drawing exactly one Ace?

Given that the first card is an Ace, what is the probability of drawing another Ace?

-

Binomial Probability Distributions

Binomial probability functions are used when a single event is repeated multiple times and only two outcomes (success/failure) are considered for each repetition.

Generally, we say that there are *n* events, each with a probability *p* for success (this means the probability of failure is 1 - *p*). The probability of *r* success is given by:

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$ ***Section 6.10 of formula booklet***

The binompdf and binomcdf functions on your calculator may also be used for binomial probability.

Ex.) Ian often has a coffee in class. Each day there is a 91% chance he will bring coffee.

a) What is the probability he will have coffee at exactly three classes this week?

b) What is the probability he will have coffee for at most two classes this week?

c) What is the probability he will have coffee for at least three classes this week?

Ex.) Brett plays a game 6 times. The probability of winning exactly 4 games is 0.24. What is the probability of winning a single game?

_

Expected value

Expected value questions rely on both probabilities and some sort of quantifiable value (often money). Consider a game where a student rolls a cubical die. If the die shows a 6, the students wins $9. If the die shows 1 or 2, the students wins $6. If the die shows a 3, 4, or 5, the student wins $1.

a) Find the expected value of a single play of the game.

b) The costs $4.75 to play. Is it a fair game?

c) Ralph plays the game 30 times. How much money do we expect him to win/lose?

The Normal Distribution

If a random variable is normally distributed, then:

- 68% of the population lies within one standard deviation of the mean
- 95% of the population lies within two standard deviations of the mean
- 99.7% of the population lies within three standard deviations of the mean

This is reflected in the area under the normal curve for each scenario.

Typically, we convert raw scores to **standard scores** (or z-scores) which measure the number of standard deviations a score is above or below the mean. Positive z-scores are above the mean, negative z-scores are below the mean. The mean itself has a z-score of 0. Standard scores are computed as follows:

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

The tables in the back of your formula booklet provide the relationship between z-scores and area to the left of that z-score. Often, a sketch will be helpful.

Your GDC provides a few useful features that save you the tedium of the tables:

SHADENORM (2nd VARS) tells you the probability that a person/variable lies between given z-scores.

INVNORM (2nd VARS) tells you the z-score that has a certain area to its left.

_

Ex.) Determine the area:

a) to the left of a z-score of 1.23

b) to the right of a z-score of -0.53

c) between the z-score of -1.10 and 0.84

d) calculate

i) $P(z \leq 1.83)$ ii) $P(z \geq 0.31)$ iii) $P(-0.54 \leq z \leq 1.45)$

Ex. 1) The heights of high school basketball players are normally distributed with a mean of 177 cm and a standard deviation of 4 cm.

a) What is the probability that a randomly selected high school basketball player is taller than 182 cm?

b) What is the probability that a randomly selected high school basketball player is between 169 cm and 185 cm?

c) In a group of 250 high school basketball players, how many do you expect to be 182 cm or shorter?

Ex. 2) A test has a mean of 75 and a standard deviation of 13. Results on the test are roughly normally distributed. 80% of test-takers are expected to pass the test. Find the passing mark.

[In this case, we may want to sketch what it is that we're looking for on a normal curve]

It looks like we want the score that has an area of _________ to its left. The z-score that corresponds to this is _________. The corresponding raw score would be:

Ex.3) The scores on an aptitude test are normally distributed with a mean of 80 and a standard deviation of 11. Find:

a) the probability that a student scores lower than 88.

b) the percentage of students scored between 65 and 88.

c) $P(x \geq 65)$

d) What mark did 90% of the students score above?

_