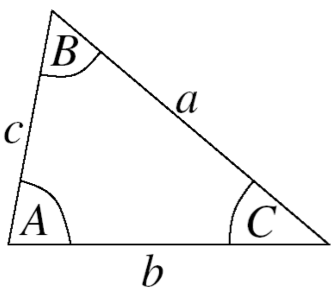
Volumes

Cylinder: (pi)(r^2)(h)
Cone: (⅓)(pi)(r^2)(h)
Curved SA of cylinder: 2(pi)(r)(h)
Curved SA of cone: 2(pi)(r)(slant height)

Sine and Cosine rule



Sine Rule

$$\frac{a}{\sin(A)} = \frac{b}{\sin(B)} = \frac{c}{\sin(C)} \quad \text{or} \quad \frac{\sin(A)}{a} = \frac{\sin(B)}{b} = \frac{\sin(C)}{c}$$

(for finding sides) (for finding angles)

Cosine Rule

$$a^2 = b^2 + c^2 - 2bc\cos(A) \quad \text{or} \quad \cos(A) = \frac{b^2 + c^2 - a^2}{2bc}$$

(for finding sides) (for finding angles)

Graph transformations:

- f(x) → f(x + 2), means to translate 2 units in the negative x-direction. So the coordinate (1, 3) → (-1, 3)
- The wording tells the truth
- f(x) → f(x/2), means to scale by factor 2 parallel to the x-axis. So the coordinate (1, 3) → (2, 3)
- Reflection is about the opposite axis (eg x replace by -x is reflection about the y-axis)

IMPORTANT Binomial:

- 100p% just means p
- Questions with a lot of conditions (eg at least, at least), break up into two layers, find the probability for the inner layer and work from there
- Binomial inception questions: Identify what your random variable in the question needs to represent: then that's your focal point. Not any other measurement
- Beware of these types of questions: "Find the probability that the 20th student she asks is the 5th student who has been infected by the disease before" → create a new random variable Y that denotes number of infected people out of 19 (because the last one must be reserved for the 5th person), find P (Y=4), then multiply by the

original binomial distribution probability that a person is infected out of the original binomial distribution n.

Identifying is it normal or binomial distribution?
- If the question you're unsure of is the last part where the previous parts were all binomial distribution question, identify if this last part requires you to find "TOTAL" something (eg money, boxes), then solve using Normal Distribution!
- If it asks you to find an individual characteristic like "a carton contains exactly 24 coupons" (where a carton has 12 family packs), then solve basic binomial distribution

Question types
X ~ B (n, p)
- Unknown p (solve with graph)
- Unknown n (must be integer, use table)
- Unknown X value (must be integer, use table)
- Given mode, solve for unknown something: X=4 is mode, P(X=4) > P(X=5), P(X=4) > P(X=3)

Normal Distribution:

- Bell-shaped curve, horizontal axis is x
- Mean in the middle, symmetrical about the mean, standard deviation sigma on both sides of the mean
- Approximately 99.7% of the data falls within **3** standard deviations about the mean (important when they ask you to draw the curve and shade the area beyond a certain value, consider this reminder to not draw the shaded area too large)
- normalcdf
- invNorm, solving invNorm with modulus: if it's I Y - 40 I > b, split into Y - 40 < -b and Y - 40 > b, and then make sure that Y is the subject eg Y < 40 - b
- Standardisation Z: change your subject X in P (X < 35), with unknown mean or variance, to subject Z. Change 35 to (35- unknown mean)/standard deviation.
- Normal distribution curve is changed such that the centre is now 0
- Expectation is the mean
- Find unknown in: don't need to use standardisation! USE GC instead (if given mew, standard deviation, and probability value alrd)

Sampling:

Definitions
- Population is the complete set of items being investigated
- Sample is a subset of the population
- Random sample of size n is a sample which is taken in such a way that every sample of size n of the population has an equal chance of being chosen. For a random sample of size n, every member of the population has an equal chance of being

chosen. Selections are independent of each other. Otherwise, sample is a non-random sample

- Random sampling is free from bias

- Actual population mean, Actual population variance
- Sample mean, Sample variance
- x bar: Unbiased estimate for population mean, s^2: Unbiased estimate for population variance
- Sample variance is not an unbiased estimate of the population variance
- Use GC to input values in a table to get mean and stuff: STATS, LIST, 1-Var. Sx is the square-rooted version of the unbiased estimate of population variance
- How to carry out random sampling: Label each item from 1 to 100, use a random number generator to generate 40 numbers to obtain a sample of 40 items

- Introduces sample mean X bar, to normal distribution
- X bar = (X1 + X2 +... Xn)/n
- Expectation does not change
- E (X bar) = mew
- Var (X bar) = sigma^2 / n
- Introduces Central Limit Theorem (n more than equal to 30): for large n, the distribution of the sample mean X bar, is approximately normal
- 2 formulas for unbiased estimate of population variance in MF27
- Important formula for UE of population variance that isn't in MF27: s^2 = n/(n-1) x sample variance
- Sample variance is (standard deviation)^2

Hypo testing:

- Tests follow normal distribution and its curve, consider CLT if relevant
- Test statistic is the small x bar you use to test. Z is the value of the standardised test statistic if there is an unknown
- The question will always tell you the sample (important if they gave the standard deviation in a sample, which means you need to find the unbiased estimate of population variance using the formula)

Steps

1. Let mew be population mean time
2. Test Ho : mew = a against H1 : mew </>/not equal to a at b% level of significance
3. If given population variance, use it. If not given it, find unbiased estimate of population variance. Find unbiased estimate of population mean if need be.
4. Under Ho, as n is large, X bar ~ N(a, popn variance/n) by CLT if necessary
5. State value of test statistic
6. Calculating p-value using GC:
    - [STATS]
    - [TESTS]
    - 1: Z-Test

    - Input mewo, popn standard deviation, small x bar, n, (left, right, or centre)
7. If p less than level of significance, reject Ho as there is sufficient evidence at the b% level of significance to conclude that (event) has/has not occurred.

- When we have unknowns in our normal distribution parameters
- Standardise using Z ~ N(0, 1)
- Critical region is where the value of test statistic lies for Ho to be rejected
- invNorm: eg 8% level of significance is your probability area under the normal distribution curve, the value on the horizontal axis is what you need to find, any value that falls in the area of 0.08 is in the critical region. invNorm to find the value on the horizontal axis that demarcates the border
- Small z value is the value on the horizontal axis, z = (X - mew)/(variance/n)^1/2
- Finding z value on the horizontal axis is the same method as finding P(X < a) unknown a using invNorm, but the area is level of significance
- A-Level Critical Region is in terms of x bar: Find formally critical region by invNorm to find z value, then find x bar. Critical region: x bar < / >

Correlation and Linear Regression:

- Scatter diagrams work with the pure data
- Sketch scatter diagrams: [STATS], [EDIT], [2nd][STATPLOTS]. Label x and y axes, label the extreme x and y values
- Product moment correlation coefficient is a measure of the linear correlation between the two variables, using the pure data [STAT], [CALC], LinReg
- r > 0 → positive linear correlation between X and Y (weak, moderate, strong)
- r < 0 → negative linear correlation
- Value of r can be close to 1 but scatter diagram may show a non-linear relationship
- Removal of outliers can dramatically change the correlation coefficient. Unit change does not affect correlation coefficient because r value has no units
- Two variables may not have a linear correlation (r close to 0), but they could have a non-linear relationship
- Linear regression is a way to model the relationship between two variables by fitting a linear equation to a set of observed data, draw scatter diagram first.
1. Linear regression line of y on x (x values as accurate, x is independent variable)
2. Linear regression line of x on y (y as the accurate, y is independent variable)
    - Deviation of point from line of best fit (residuals)
    - Residuals are squared because residuals can be negative/positive, adding the errors together may cancel each other out. If sum of the squares of the residuals is used, representations of errors are all made positive, allowing a more accurate estimate of least squares regression line

- GC regression line: 3rd bulletpoint

Sampling and testing

Unbiased estimate of population variance:

$$s^2 = \frac{n}{n-1}\left(\frac{\Sigma(x-\bar{x})^2}{n}\right) = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$$

Regression and correlation

BLUE : reg l of y on x
GREEN: reg l of x on y

Estimated product moment correlation coefficient:

but without the sq root

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\left\{\Sigma(x-\bar{x})^2\right\}\left\{\Sigma(y-\bar{y})^2\right\}}} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)\left(\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right)}}$$

Estimated regression line of y on x:

also equal to

$$y - \bar{y} = b(x - \bar{x}), \quad \text{where } b = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

but without the square root

$\Sigma(y-\bar{y})^2$

$x - \bar{x} = d(y-\bar{y})$

WILCOXON SIGNED RANK TEST

P is the sum of the ranks corresponding to the positive differences,

- Q is the sum of the ranks corresponding to the negative differences,

- GC sketching regression line of y on x: XList → L1 (independent), YList → L2
- GC sketching regression line of x on y: Same set of data, XList → L2, YList → L1
- Finding values of x bar and y bar: [STAT], [CALC], 2: 2-Var Stats, L1 into XList, L2 into YList because both regression lines intersect at (x bar, y bar)
- Sketching both regression lines, put both equations into normal graph page (make y the subject for regression line of x on y before putting it in)
- If scatter diagram shows that relationship between y and x is non-linear, it is not suitable to find linear regression line for y on x
- Change the model of the x y equation to a linear one (linearize it), change GC [STATS] [EDIT] (eg ln x). Find a regression line now that non-linear data points have been linearized.
- If given multiple models to test, find the one with the best r value by subbing in the different models into GC. r very close to 1 indicates strong positive linear correlation.
- Comparing between models, look at whether or not the shape of the given models corresponds to the shape of data points. As x increases, y increases/decreases at an increasing/decreasing rate, thus the model is suitable/not suitable…