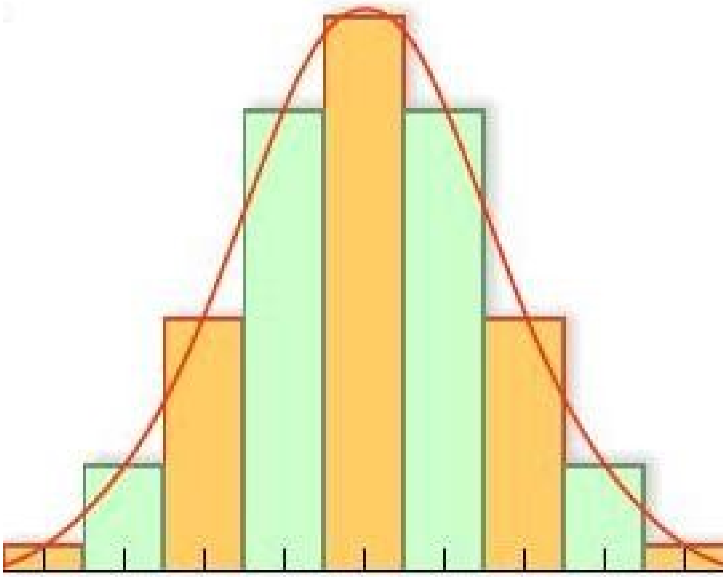
"Do-and-don't" for S2 statistics

S2 binomial, Poisson, normal, uniform



DO: # Know discrete is |||||| and continuous is ________

 # Know binomial is discrete and like number of heads when tossing coins

 # Know Poisson is discrete and like number of goals when playing football

 # Know normal is continuous and like errors in a manufacturing process

 # Know the continuous uniform distribution is continuous and like waiting time for a train which goes exactly every 20 minutes

 # Know binomial → Poisson as n becomes large with p fixed and small (<0.05)

 # Know binomial → normal as n becomes large (for any p, but especially middling p)

 # Know Poisson → normal as λ becomes large

 # When using normal approx to binomial and Poisson, do continuity correction: P(binom or Poisson variable is some whole number values) ≅ P(normal variable is any number which rounds to those whole number values)

 # Learn by heart the conditions for when to use distributions
http://www.memrise.com/course/336931/a-level-maths-edexcel-s2/

DON'T # Fail to make continuity correction

DO # Think: P(binom or Poisson variable is some whole number values) ≅ P(normal variable is any number which *rounds to* those whole number values)

DON'T # Look up or calculate $P(X=n)$ when you want $P(X\leq n)$, or vice versa

DO # Always write out the P you're looking up or calculating, $P(X\leq n)$; or $P(X=n)$; or $P(X>n)$, which is $1-P(X\leq n)$; etc.

DON'T # Write e.g. $X\sim B(100,0.05)$ without saying what X is

DO # When you use approximations, be clear, e.g.:
$X\sim B(100,0.05)\approx Po(5)$ or $X\sim B(100,0.2)\approx N(50,40)$ or $X\sim Po(50)\approx N(50,50)$

DO: # State H_0 = *null hypothesis* = default belief in absence of new evidence, e.g. coin is unbiased, medicine makes no difference

DO: # State H_1 = *alternative hypothesis* = suggestion to be checked, e.g. coin is biased, medicine helps cure you

DO: # State H_1 as *one-tailed*, e.g. test whether coin biased to heads alone, or *two-tailed*, e.g. coin biased either way

DO: # State *significance level* = your standard of how improbable a result has to be on the assumption of H_0 for you to reject H_0 and go for H_1

DO: # State *critical region* = range of results (one-tailed or two-tailed) which is so far out from what's expected with H_0 that probability < significance level

DO: # State *actual significance level* = probability of result being in critical region (i.e. less than pre-set significance level)

DO: # *Draw a picture* of the range of possible results and the critical region, double-checking on one tail or two

DO # Learn definitions about sampling by heart: http://www.memrise.com/course/336931/a-level-maths-edexcel-s2/.

DON'T # Mix up H_0 and H_1.

DO # Remember H_0 is what you would go for with 0 (zero) new evidence

DON'T # Get continuity corrections wrong if you're using Normal approximation

DO # Think: P(binom or Poisson variable is some whole number values) = P(normal variable is any number which *rounds to* those whole number values)

DON'T # Think or write that a result *proves* H_0 (or H_1). Think, and write: "The evidence is not enough to reject H_0" or "The evidence is enough to reject H_0". And write what it means: "The evidence is enough to suggest the coin is biased", or "the evidence is not enough to suggest the coin is biased".

DO: # calculate E(X)= $\sum x_i p(x_i)$ **or =** $\int x f(x) dx$

DO: # calculate E(X²)= $\sum x_i^2 p(x_i)$ **or =** $\int x^2 f(x) dx$

DO: # calculate var(x)=(E(X))²–(E(X))²

DO: # Remember probability density function f(x) does not measure a probability. For a continuous random variable, p(any given value)=0. Any function f(x) can be a pdf if f(x)≥0 for all x and $\int_{-\infty}^{\infty} f(x) dx = 1$

DO: # Calculate cumulative distribution function

F(x)= $\int_{-\infty}^{x} f(t) dt$ **is a probability. F must be continuous, always increasing, and → 1 as x→∞**

DO # Calculate median by F(m)=0.5

DO # Calculate skew as the opposite of what you'd think: if the graph is tilted to the right, that's negative skew

DON'T # Fail to show values of F for the whole range (usually F is 0 for a whole range of low x, and 1 for a whole range of high x)

DON'T # Fail to "fit together" the values of F at the points where different functions defining it join.

DO # When you introduce each random variable X, say what it is, e.g. X=number of heads from 20 tosses of coin.

DO # Use capitals for random variables.