S1 Cheat Sheet

<table>
<tr><td colspan="3">Using Your Calculator</td></tr>
<tr>
<td>You can handily use your calculator's STAT mode to calculate the mean and variance of a frequency table, find r (PMCC) and the a and b for the line of best fit, along with a number of summations such as $\Sigma x, \Sigma xy, \Sigma x^2, \ldots$

Setting Up:

Press SHIFT → MODE for setup. Press down, select STAT then turn frequency mode On (your calculator will preserve this setting even when turned off).</td>
<td>Getting Started

Press MODE → STAT. "1-Var" means 'one variable'. Use "$a + bx$" for two variables (e.g. regression and correlation). Note that frequency is not a variable.

Entering Data

Press = after each number is entered. Use the arrow keys to navigate around the table. Press AC once finished to go to calculation mode.</td>
<td>Once in calculation mode, if you want some statistical expression in your calculation, e.g. $\bar{x}$ or r or Σx, press SHIFT → 1 (for STAT).

Var: Contains $\bar{x}, n$ (same as Σf), σ_x

Sum: $\Sigma x, \Sigma xy, \Sigma x^2, \Sigma y^2$

Reg(ression): a, b, r (using for regression and correlation chapters).</td>
</tr>
<tr><td colspan="3">Exam Note: Suppose you need to calculate S_{xx}. It is fine to use your calculator to get summations like Σx and so on. But you must still show the substitution of the numbers into the relevant formula. The calculator is incredibly handy for checking the value of $\sigma, a, b, r, \bar{x}$, but should be used to check, not to do the question for you.
See http://www.drfrostmaths.com/resource.php?id=11591 for a Virtual Silver Calculator giving more detail on the buttons and modes available.</td></tr>
</table>

Chapter	Usual types of questions	Tips	What can go ugly
1 – Mathematical Models	• "What are the stages of a mathematical model?" • "Write down reasons for using mathematical models."	• Stages are (memorise!) 1. Recognition of a real-world problem. 2. Statistical model is devised. 3. Model used to make predictions. 4. Experimental data is collected. 5. Comparisons are made against the devised model. 6. Statistical concepts are used to test how well the model describes the real-world problem. 7. Model is refined. • Reasons for using statistical models (memorise!) 1. Simplify a real-world problem. 2. To improve understanding/describe/analyse a real-world problem. 3. Quicker and cheaper than using real thing. 4. Predict possible future outcomes. 5. Refine model/change parameters possible.	• Not much provided you memorise what's on the left! Such exam questions are rare but students have been thrown because they hadn't expected to be tested on Chapter 1, or their teacher has skipped the chapter.
2/3 – Measures of Location and Dispersion	• Calculating mean and standard deviation from a grouped frequency table. • Calculating quartiles and the median using interpolation.	• Calculating $\bar{x}$ is two marks. The method mark is for the division, NOT showing your calculation of midpoints times frequency. Writing out the calculation in full takes too much time and prone to calculator-punching errors. Use STAT mode to get Σx, then just show $\frac{\Sigma x}{n}$ as your calculation (i.e. with the Σx and n subbed in), followed by the final answer. Do not round your mean or give as a fraction! (the latter which can be penalised) Note that for grouped data Σx on your STAT mode is Σfx and n gives Σf.	So much can go wrong! From experience, students: • misremember the standard deviation formula (either forgetting to square the mean, or forgetting to square root the variance to

- Stating the effects of coding on quartiles/median /mean/standard deviation /interquartile range.
- Be able to combine means.

- Mmemonic for **variance**: "**The mean of the squares minus the square of the mean (msmsm**)"
    1. Ungrouped data: $\sigma^2 = \frac{\Sigma x^2}{n} - \bar{x}^2$
    2. Grouped data: $\sigma^2 = \frac{\Sigma f x^2}{\Sigma f} - \bar{x}^2$

  Don't forget to square root for variance! And use your exact value for $\bar{x}$. Note that Σx^2 on your calculator is effectively $\Sigma f x^2$ when frequencies are given.
  Standard deviation can roughly be thought of as "the average distance of values from the mean" (e.g. if you just had 3 and 7 as values, the standard deviation IS 2 because they're both 2 away from the mean of 5). Thus do a common-sense check that your value looks sensible.
- **Coding:** Note that you'd always calculate statistics like the mean on the *coded* data, not the original data. The idea is then you work out what this stat would be on the original data.
  e.g. If coding is $y = x - 2$, where original variable is x and new coded one y, we'd calculate say $\bar{y}$, then add 2 to get the original mean $\bar{x}$. Make sure therefore that you check whether the question wants you to work forwards or backwards.
    1. Mean/median/quartiles: Affected by all of $\times, \div, +, -$
    2. Standard deviation/range/interquartile range: Only affected by $\times, \div$
    3. Variance: Any scale factors will be squared. e.g. If $y = 3x + 2$, $\sigma_y^2 = 9\sigma_x^2$
- To work out the item number required for say the lower quartile:
    1. For listed data: Find $\frac{n}{4}$ (similarly $\frac{n}{2}$ for median, or $0.34n$ for 34th percentile, etc.)
       If a decimal, round up. If whole, use halfway between this item and one after.
    2. For grouped data: Find $\frac{n}{4}$. DO NOT ROUND OR ADJUST. Then use linear interpolation.
- **Linear Interpolation**
  Be vigilant of gaps in the intervals vs no gaps. If gaps, just expand the intervals first so fill the space. So $3 - 5$ would become $2.5 - 5.5$. Note this increases the class interval by 1.

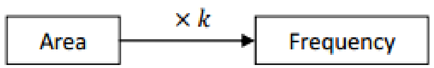
Weights of owls w (kg)	$3 - 5$	$6 - 8$	$9 - 13$
Frequency	3	4	5
Cumulative Frequency	3	7	12

  Suppose we want median. $\frac{12}{2} = 6$ so use 6th item (this is grouped data so don't do anything to the 6). Construct a suitable line. I put units on the interval values to avoid confusion between frequencies and values of the variable (in this case weight).

  3 ——————————— 6 ——— 7
  5.5kg ——————— ? ——— 8.5kg

  Since 6 by observation is $\frac{3}{4}$ of the way along the interval frequency-wise, we want to go $\frac{3}{4}$ of the way along the class interval. Thus:
  $Q_2 = 5.5 + \left(\frac{3}{4} \times 3\right) = 7.75kg$
- Combined means: $\bar{x} = \frac{n_1\bar{x_1} + n_2\bar{x_2}}{n_1 + n_2}$ (General principle is to add the sums of the values and divide by the combined number of things there are)

get the standard deviation).
- don't observe whether there are 'gaps' in the intervals and thus are penalised heavily for linear interpolation.
- introduce rounding errors because they don't use their exact mean in the standard deviation calculation.
- think that Σx^2 means $(\Sigma x)^2$.
- round the item number for linear interpolation. If $n = 40$ and you're interpolating to find the LQ, use the 10th item: don't do ANY adjustment! Similarly, if $n = 35$, use 8.75, not 9!

4 – Continuous Random Variables	• Constructing a box plot (probably with outliers). • Construct a stem-and-leaf diagram (or back-to-back stem-an-leaf). • Calculate interquartile range/median from a stem-and-leaf diagram. • Find widths/heights of bars in a histogram. • Find the frequency of a interval using a histogram (e.g. number of runners with a time under 10s). • Use histogram to complete frequency table and vice-versa. • Use frequency table formed from histogram to make subsequent Chp2/3 calculations, e.g. mean. • Calculate and interpret skew.	• There's 3 ways you can find skew: 1. For histograms/probability distributions, just observe the **shape**. If the 'tail' is in the positive direction, then positive skew. 2. Using **quartiles**: $Q_3 - Q_2 > Q_2 - Q_1$ means right box in box plot is larger, so positive skew. 3. If positive skew then $mean > median$. I remember as the mean determining the skew; e.g. large salaries in the positive tail drag mean up but not median, so positive skew. It is vital you use whatever data is available from the previous parts of the question. If you've just worked out the quartiles but don't know the mean, then it would be stupid to give your justification of the skew involving mean and median. • For box plots, definition of outlier boundary is always given, but generally: $LQ - 1.5 \times IQR$ and $UQ + 1.5 \times IQR$ You **must** explicitly show the calculation you used to calculate your boundaries, and there's marks associated with this. Outliers are indicated using × symbol. It's possible to have multiple outliers at each end (or none). • There's two possibilities for the end points of the whiskers when there's an outlier on that end, and mark schemes accept both: either use the outlier boundary itself, or the smallest/greatest value which is not an outlier (I prefer the latter). But be consistent in which you use. • For **histogram** questions involving finding width and height of bars, just find the scale factor between class width and actual width, and between frequency density and actual height. e.g. If a bar with class width 3 and frequency density 10 has 'drawn' width 6cm and height 5cm, then the scale factors are 2 and 0.5 respectively. Then if some other bar has class width 5 and frequency density 8, then its width and height are 10cm and 4cm. • Remember that in histograms, area is not necessarily equal to frequency, they are proportional.  Key is to use a known area with known frequency to find the scale factor k. This scale factor can then be used to convert other areas in the histogram into frequencies. (See my Chp4 slides) • As per Chps2/3, you must check whether the intervals have 'gaps', and adjust appropriately when calculating frequency density. $frequency\ density = \frac{frequency}{class\ width}$ • When asked to find the mean, median, quartiles or variance of a histogram, first use the histogram to generate a grouped frequency table. Then use this table as you usually would to calculate these statistics • "Explain why a histogram is appropriate for this data" → "Data is **continuous**." (No credit given for talking about different possible class widths) • Possible wordy questions: "Explain whether the mean or median would be more suitable for this data." "The median because the data is skewed". The question wouldn't be asked if the data wasn't skewed. The logic is that extreme values affect the mean but not the median, and thus the median may be more suitable.	• Not showing your outlier boundary calculations for box plots. • Referring to things you haven't calculated (e.g. quartiles) when justifying what the skew of your data is. • Not considering gaps (if present) for histograms. • Just being careless when finding areas within a histogram (e.g. misreading scales). • Getting positive and negative skew the wrong way round.

5 – Probability

- Find probabilities by enumerating outcomes.
- Use Venn diagrams to find frequencies.
- Use Venn diagrams to find unknown probabilities.
- Use the laws of probabilities to calculate probabilities, e.g. involving conditional probabilities, independent and mutually exclusive events, etc.
- Use tree diagrams to determine probabilities.
- Be able to construct a Venn Diagram that takes into account the relationship between events.

- One neglected type of question, usually found at the end of Discrete Random Variable exam questions, is to be able to enumerate different matching outcomes so we can find the total probability.
  Suppose we had the following distribution:

S	1	2	3	4
$P(S)$	0.1	0.2	0.3	0.4

  Then if X is defined as the product of two spins of this spinner, we might be interested in finding $P(X > 2)$. As per GCSE, we could list the possible outcomes, find the probability of each, then add them together. You can be clever however in how you combine possibilities:

S_1	S_2	$P(S_1 \times 2_2)$
1	≥ 3	$0.1 \times 0.7 = 0.07$
2	≥ 2	$0.2 \times 0.9 = 0.18$
≥ 3	any	$0.7 \times 1 = 0.7$

Thus $P(X > 2) = 0.07 + 0.18 + 0.7 = 0.95$. See Ex1 in my slides: http://www.drfrostmaths.com/resource.php?id=11650

- Venn Diagrams with frequencies are fairly self-explanatory. Just be on the lookout for the word 'given that' when asked to find a probability. e.g. If "given that they are cyclist", then your probability is out of the cyclist set/circle, rather than out of everyone.
- Ensure you can correctly identify regions in a Venn Diagram involving intersections, unions and complements, e.g. $A \cap B'$. Recognise that $(A \cup B)'$ is the same as $A' \cap B'$.
- When asked to find probabilities given other probabilities, Venn Diagrams are particularly a good strategy if probabilities involve intersections and complements, e.g. if you know $P(A)$ and $P(A \cap B)$ and need to find $P(A \cap B')$, then via a Venn Diagram it would be much easier to see that $P(A \cap B') = P(A) - P(A \cap B)$.

All Laws of Probability:

- **If A and B are mutually exclusive (i.e. "can't happen at the same time") then:**
  $P(A \cap B) = 0$
  $P(A \cup B) = P(A) + P(B)$
- **If A and B are independent (i.e. "one event does not influence the other") then:**
  $P(A \cap B) = P(A)P(B)$
  $P(A|B) = P(A)$
- **In general:**
  $P(A|B) = \frac{P(A \cap B)}{P(B)}$ (I remember this as 'the probability of the intersection divided by the probability of the thing you're conditioning on')
  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (the Addition Rule)

- Nonsensical notation, e.g. $P(0.3)$, when you mean that the probability is 0.3.
- Using $P(A \cap B) = P(A)P(B)$ when you weren't told the events were independent, or similarly $P(A \cup B) = P(A) + P(B)$ when you weren't told they were mutually exclusive.
- Just generally getting stuck and not knowing how to proceed to calculate a probability given other ones. As per the advice, try drawing a Venn Diagram. Were you told any relationships between the events? (i.e. mutually exclusive or independent)
- Accidentally doing $P(A|B) = \frac{P(A \cap B)}{P(A)}$
- Writing your conditional probability wrong in the first place, e.g. if "Given that A, find probability of B", then $P(A|B)$ would be wrong. Remember the "|" symbol means "given that", so check by reading the symbols out to yourself.

		• There are certain reactionary things you should write out immediately after reading the initial blurb of a probability question, before even starting part (a) of the question: 1. If you see that "A and B are independent", immediately write out $P(A \cap B) = P(A)P(B)$, as you might forget you can use this otherwise. 2. If you see a conditional probability given, write out the conditional probability formula involving those events. • Always be on the lookout for the words 'given that'. If so, you'll have a conditional probability and hence need to use the appropriate formula. Tends to come up at end of question. • Do you have a mixture of $P(A \cap B)$, $P(A \cup B)$, $P(A)$ and $P(B)$ in the problem? Write out the Addition Rule and sub things in. There's harder 4 mark questions that involve use of this. • You are sometimes asked to construct a Venn Diagram which encapsulates the relationship between events. Remember that if events are mutually exclusive, the circles must be separate. But if you're not told they're mutually exclusive, you must assume they can overlap. Independence does not affect the Venn Diagram. • Suppose we had a Tree Diagram where we had event A followed by event B followed by C. 1. $P(C)$? As per GCSE, find all paths which match and each probabilities for each path. $P(C) = P(A \cap B \cap C) + P(A \cap B' \cap C) + P(A' \cap B \cap C) + P(A' \cap B' \cap C)$ 2. $P(B)$? Note we need not consider C as it happened after B. $P(B) = P(A \cap B) + P(A' \cap B)$ 3. $P(C\|B)$? Harder tree questions ask you to calculate some conditional probability. Just use the conditional probability formula as normal. Students tend to find probability questions the hardest. The key is just practising the four different types of questions. See my slides http://www.drfrostmaths.com/resource.php?id=11650	
6 – Correlation	• Calculate S_{xx}, S_{xy}, S_{yy} • Calculate the PMCC, r • Interpret a correlation coefficient. • State the effect of coding on PMCC. • Comment on whether a given claim about correlation is justified.	• Ensure you know what your x variable is and what your y variable is, if the question uses different letters. The first row in any table given is always the x. • "Interpret the correlation coefficient". 'Interpret' means a non-statistical wordy interpretation is required, e.g. "As the altitude increases, the temperature increases". • Generally >0.7 and <-0.7 is considered strong correlation. • "Is their claim justified?". Generally two marks. e.g. "Claim justified (1 mark) as the value of r is close to 1 (1 mark)". • All linear codings (adding/multiplying/dividing/subtracting) have <u>no effect</u> on PMCC. • S_{xx} is affected by coding in the same way as variance is. So if we used $t = 3x + 1$, then $S_{tt} = 9 \times S_{xx}$ • Use the STAT mode to check your value of r if the original data is given. You obviously still need to show your calculations for S_{xx} and so on.	• Forgetting to square root denominator in formula for r. • Stating the correlation (e.g. "negative correlation") when you were asked to **interpret** it.

<table>
<tr>
<td>7 – Regression</td>
<td>
• Given data or summarised data, find the line of best fit $y = a + bx$
• Comment whether a prediction made using your $y = a + bx$ formula is reliable.
• Find the equation of the line of best fit using the original non-coded variables.
• Be able to ‘interpret’ the values of a and b.
• Justify which is the explanatory variable.
• Explain why the data justifies fitting a regression line.
</td>
<td>
• $y = a + bx$ then $b = \frac{S_{xy}}{S_{xx}}$ and $a = \bar{y} - b\bar{x}$. Use your exact value of b in your calculation for a; avoid rounding errors! If the question uses variable letters other than x and y, match up the variables first so you don’t get S_{xx} and S_{yy} mixed up. The first row is always x.
• Remember that your calculator’s STAT mode can check your values of a and b if the original data is provided. You’d be a donut not to check.
• Worded questions:
1. “Explain why this diagram would support the fitting of a regression line of y onto x.”
→ “The points are close to an implied straight line of best fit.”
2. “<u>Interpret</u> the gradient b”
→ “As [the altitude] <u>increases by 1</u>, [the temperature] increases/decreases by ___” (This needs to match the context)
3. “Interpret the y-intercept a”
→ “[The temperature] is a when [the altitude] <u>is 0</u>”
4. “Which is the explanatory variable. Explain your answer.”
→ “[The altitude] is the explanatory variable as [the altitude] influences [the temperature]”
• Reliability questions: (After using your equation $y = a + bx$ to estimate a y value)
“Comment on the reliability of your estimate.” One of:
1. “Reliable (1 mark) as value inside the range of the data (1 mark)”
2. “Unreliable (1 mark) as value outside the range of the data (1 mark)”
3. “Reliable (1 mark) as value just outside the range of the data (1 mark)”
• Coding is simple; simply sub in the expressions and rearrange.
e.g. If $y = 3 + 2x$ and you had coding $x = 3a + 4$ and $y = b - 1$, then:
$$b - 1 = 3 + 2(3a + 4)$$
$$b = 12 + 6a$$
</td>
<td>
• Mixing up your x and y variables (particularly when they use different variable letters, and hence incorrectly using $b = \frac{S_{xy}}{S_{yy}}$ instead of $b = \frac{S_{xy}}{S_{xx}}$
• Forgetting to write out the equation of the regression line after working out a and b.
• Accidentally mixing up a and b. Remember that $y = a + bx$. You may be used to $y = mx + c$ from GCSE, where the x term generally comes first. In stats, the x terms comes second.
</td>
</tr>
<tr>
<td>8 – Discrete Random Variables</td>
<td>
• Find a constant k using a probability distribution/function.
• Find a constant k using a cumulative distribution function.
• Turn a probability distribution into a cumulative distribution, and vice versa.
• Calculate the expected value or variance of a discrete random variable.
• Find two missing probabilities (using
</td>
<td>
• Note the difference between a probability function and a probability distribution:
Example probability function: $p(x) = \begin{cases} 0.1x, & x = 1,2,3,4 \\ 0, & otherwise \end{cases}$
Equivalent probability distribution:
<table>
<tr><td>x</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>$p(x)$</td><td>0.1</td><td>0.2</td><td>0.3</td><td>0.4</td></tr>
</table>
• Note the difference between shorthand $p(x)$ (with lowercase p) and longhand $P(X = x)$ (with capital P).
• Recall that $F(x) = P(X \leq x)$ i.e. the running total of the probability. If given a decimal value (and outcomes are integers), then round down, i.e. $F(3.7) = F(3)$.
• When asked to find a missing constant (say k) in a probability function, first find the probability of each outcome (i.e. find the probability distribution), then use fact that probabilities sum to 1.
• When similarly asked the same of cumulative distribution function, use fact that $F(\ldots)$, for the last outcome is 1.
</td>
<td>
• When calculating $E(X^2)$ for use in $Var(X)$ formula, remember that the outcomes only are squared, not the probabilities.
• Common error:
$Var(3X) = 3Var(X)$
</td>
</tr>
</table>

	simultaneous equations) when the expected value is given. • Calculate $E(X^2)$. • Find the probability of an inequality involving a discrete random variable.	• If given an inequality within a $P(\dots)$, solve the inequality first. e.g. $P(2X+1<7)=P(X<3)$ • $E(X)=\Sigma x\, p(x)$ and $Var(X)=E(X^2)-E(X)^2$ (the "msmsm" mnemonic still applies here) • If you have two missing probabilities in a probability distribution, and are given the expected value, use: (a) Fact that probabilities add up to 1 and (b) Equation for expected value. This gives you two simultaneous equations in terms of these variables. • Note that rearranging the expected value equation gives $E(X^2)=Var(Y)+E(Y)^2$. On two occasions in the past, exam questions have asked for $E(X^2)$. • Coding: Same rules apply as with other chapters. 1. For $E(\dots)$, tip is to rub out E then write it back round X, e.g. $E(3X+1)=3E(X)+1$ 2. For $Var(\dots)$, remember scale factors are squared, and + and – have no effect. e.g. $Var(3X+1)=9Var(X)$ and $Var\left(\frac{1}{2}X-1\right)=\frac{1}{4}Var(X)$ • Don't forget that discrete random variable questions can have a second half involving probability questions more like Chapter 5. See notes on this section.	
9 – Normal Distribution	• Find the probability of a value being more/less than a value. • Find the probability of a value being in a two-ended range, i.e. $P(a<X<b)$ • Find the value that gave rise to a probability. • Find the value in a double-ended range, e.g. a such that $P(\mu<X<\mu+a)=0.3$ or $P(\mu-a<X<\mu+a)=0.4$ • Find conditional probabilities involving the normal distribution. • Find a missing μ and/or σ.	For most of the examples below we will use the example of IQ, which by definition has the distribution $X\sim N(100,15^2)$ **Some core stuff first:** • z represents **the number of standard deviations above the mean**. So an IQ of $x=130$ corresponds to $z=2$ given it's two standard deviations above the mean of 100. Similarly $x=85$ gives $z=-1$. It's therefore not difficult to see therefore why $Z=\frac{X-\mu}{\sigma}$ • The z-table allows you to find the probability **up to** a given z-value. The first **table requires the inequality to be <, the z-value to be positive and the probability to be at least 0.5**. • If in $P(Z<z)$ you either have to change the direction of the inequality or the sign of z, you do 1 minus. If you change both, the "1 minuses" cancel each other out. e.g. $P(Z<2)\rightarrow\ 0.9772$ $P(Z<-2)=1-P(Z<2)=0.0228$ $P(Z<-2)=1-P(Z<2)=0.0228$ $P(Z>-2)=P(Z<2)=0.9772$ • The same applies if using the z-table backwards. If using the first table, you require the probability to be at least 0.5 and the inequality to be <. $P(Z<z)=0.58\ \rightarrow 0.2$ $P(Z<z)=0.42\ \rightarrow\ P(Z<-z)=0.58\ \rightarrow -z=0.2\ \rightarrow\ z=-0.2$ $P(Z>z)=0.58\ \rightarrow\ P(Z<-z)=0.58\ \rightarrow -z=0.2\ \rightarrow\ z=-0.2$ (note in the last example above we had to change both the direction and the sign to ensure 0.58 stays above 0.5) $P(Z>z)=0.42\ \rightarrow\ P(Z<z)=0.58\ \rightarrow\ 0.2$ • The approach above is the 'super safe' but slightly cumbersome approach. Easier is to think whether we're below the mean or above the mean. If you know you're below (e.g. you know you're in the bottom 30%), the z value is negative, otherwise positive. • You MUST use the value from the second z-table if you have a 'nice' probability. e.g. $P(Z<z)=0.2\ \rightarrow\ z=-0.8416$ (as explained above, we're in the bottom 20% so we know we're in the bottom half)	• Getting the sign of your z value wrong. • Being sloppy with rearrangement and doing say $\frac{10}{\sigma}=20$ and doing $\sigma=\frac{20}{10}$ • As with probability, just writing nonsensible notation, e.g. using $P(Z<0.6)$ when 0.6 is the probability. • Forgetting to standardise. • Not using the second z-table when you are able to do so (this costs you at least a mark).

- **Questions involving z-table being used forwards.**
  "Find the probability that someone has an IQ above 130."
  $P(X > 130)$ (Express question probabilistically)
  $= P\left(Z > \frac{130-100}{15}\right) = P(Z > 2)$ (Standardise using $Z = \frac{X-\mu}{\sigma}$)
  $= 1 - P(Z < 2)$ (Manipulate so we can use table)
  $= 0.0228$ (Use table)

  "Find the probability that someone has an IQ between 85 and 130."
  $P(85 < X < 130)$
  $= P(X < 130) - P(X < 85)$ (in general $P(a < X < b) = P(X < b) - P(X < a)$)
  $= P(Z < 2) - P(Z < -1)$
  $= P(Z < 2) - (1 - P(Z < 1))$ …

- **Questions involving z-table being used backwards.**
  "70% of people have an IQ above you. What is your IQ?"

  Quick way:
  $P(X > x) = 0.7$
  $\frac{x-100}{15} = -0.5244$ (as we're in the bottom half)
  $x = 92.134$

  Slow but safer way:
  $P(X > x) = 0.7$
  $P(Z > z) = 0.7$ (standardise)
  $P(Z < -z) = 0.7$
  $z = -0.5244$ $\frac{x-100}{15} = -0.5244$
  $x = 92.134$

  (I will do the 'quick' way from now on)

  "Find the value of a such that $P(100 < X < a) = 0.3$"
  Key is to convert two-ended range into just one-ended. With a quick sketch it's clear that
  $$P(X < a) = 0.8$$
  (Alternatively we could have done $P(100 < X < a) = P(X < a) - P(X < 100) = 0.3$
  $$P(X < a) - 0.5 = 0.3 \rightarrow P(X < a) = 0.8$$
  We can then proceed in the usual way.

  "Find the value of a such that $P(\mu - a < X < \mu + a) = 0.3$
  With a quick sketch we can see we have an area symmetrical about the mean. If 0.3 is in the middle then we have 0.35 at each tail. Thus $P(X < \mu + a) = 0.65$
  $$\frac{\mu + a - \mu}{15} = \frac{a}{15} = 0.39 \rightarrow a = 5.85$$

- **Questions involving a missing μ and/or σ**

  <u>"If $X \sim N(70, \sigma^2)$ and $P(X < 80) = 0.6$, find the value of σ."</u>

  Method is pretty much exactly as before:

$$\frac{80-70}{\sigma} = 0.2533$$
$$\sigma = \frac{10}{0.2533} = 39.48$$

  "If $X \sim N(\mu, \sigma^2)$ and $P(X > 20) = 0.3$ and $P(X < 10) = 0.25$, find μ and σ."

$$\frac{20-\mu}{\sigma} = 0.5244 \qquad \frac{10-\mu}{\sigma} = -0.67$$
$$20 - \mu = 0.5244\sigma \qquad 10 - \mu = -0.67\sigma$$

  Then subtracting: (being very careful about your negatives)

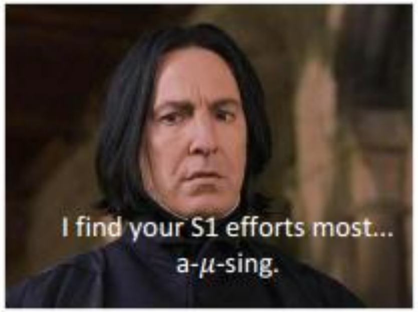
$$10 = 1.1944\sigma$$
$$\sigma = 8.3724$$
$$\mu = 20 - 0.5244 \times 8.3724 = 15.6095$$

- <u>Questions involving conditional probabilities.</u>

  $X \sim N(100, 15^2)$. Given that someone's IQ is over 85, find the probability their IQ is over 115.

$$P(X > 115\ |X > 85) = \frac{P(X > 115 \cap X > 85)}{P(X > 85)} = \frac{P(X > 115)}{P(X > 85)}$$
$$= \frac{P(Z > 1)}{P(Z > -1)} = \frac{0.1587}{0.8413} = 0.1886$$

  Note that $X > 115 \cap X > 85$ is equivalent to $X > 115$, as $X > 85$ is already implied by $X > 115$. I call these "redundant events". They come up quite frequently in S1 exams.



Harry Potter stats puns (c) Lord Voldermort.
Copyright infringement punishable by Cruciatus.