Formulas and Notation:
- Use formula card as a study guide! At this point you need to know/understand everything on the front of the first page, the confidence interval/sample size information in the Population Proportion section on the second page, and how to read the Table A1.
- The information below uses the layout of the formula card as a rough outline.

Standard Deviation:
- Interpreted as ‘**roughly the average distance** of the *observations* from the *mean*’.
- This does NOT involve the empirical rule for data, which only holds for bell-shaped distributions.

Probability Rules:
- Conditional probabilities
  - Finding the probability of one event happening when (or “if” or “given”) you already know the outcome of another event
  - Notation: the information that is given (i.e. known in advance) is always placed to the right of the bar |  
- Partition Rule (not shown on the yellow card, but uses conditional probabilities)
  - Takes P(B | A) and ‘flips’ it around to P(A | B) by plugging the general multiplication rule into the general definition of conditional probability together (try this yourself!)
  - This is what you do in a tree diagram problem
- To test if two events are independent
  - Check: P(A|B) = P(A)  
    i.e. knowing the outcome of event B has no effect on the probability of event A  
  - Or check: P(A and B) = P(A)P(B)
  - To save time, use the rule that is most convenient for the problem you are working on (i.e., try to make use out of any probabilities already calculated in the problem).
- Independence and mutually exclusive are two different concepts. Mutually exclusive events have no elements in common; independent events must follow one of the rules above.

Random Variables:
- A discrete random variable can take one of a countable list of distinct values (like binomial dist)  
- A continuous random variable can take any value in an interval or collection of intervals (normal, uniform, etc)

Binomial formula:
- Conditions for a binomial trial:  
  1. Two outcomes, classified as ‘success’ and ‘failure’  
  2. Fixed number of trials (n)  
  3. Independent trials (this condition is met if you have a random sample)  
  4. Same probability of success (p) for each trial.

For the interpretations: note that words that appear in bold are key; words that are italics are “blanks” that need to be filled in based on the context of the problem.
- Distribution: Binomial(n,p) ⇔ B(n,p)
- To find probabilities: for small sample sizes, use the Binomial Formula
  - Note that this formula is only for strict equality: P(X=k). If you want to find a range of values (for example: P(X ≤ k)) you must use the formula once for each value and add all the values together: for example
    - P(X ≤ 2) = P(X=0) + P(X=1) + P(X=2) but P(X<2) = P(X=0) + P(X=1)
- To find probabilities: for large sample sizes, use the Normal Approx. to the Binomial below

Normal random variables:
- Problem will usually say that a random variable is normally distributed (or “Bell-Shaped”)
- Distribution: Normal(μ,σ) ⇔ N(μ,σ)
- To find probabilities: Use Table A1
  - Table A1 gives you the area that corresponds to P(Z ≤ z*). To find the area that corresponds to P(Z ≥ z*), you must subtract the number given by the table from 1.
- It is a VERY good idea to draw pictures of the area you are looking for, or are given.
- Table A1 includes two decimal places for z*, so you should also round z* to two decimal places.
- Be careful with notation – don’t equate two numbers that are not equal!

Normal Approximation to the Binomial Distribution:
- Conditions for binomial trial are met, but n is large enough (both np ≥ 10 and n(1-p) ≥ 10) that using the Binomial formula would be too difficult.
- Distribution: approximately Normal(np, √np(1−p))
- To find probabilities: for large sample sizes, compute the z-score by plugging μ=np and σ=√np(1−p) into the z-score equation, then use Table A1 as usual

Sample Proportions:
When asked to find probabilities based on the sample proportion \( \hat{p} \):
For large sample sizes (both np ≥ 10 and n(1-p) ≥ 10)
- Distribution: approximately Normal(p, \( \sqrt{\frac{p(1-p)}{n}} \) )
- To find probabilities: compute the z-score by plugging μ=p and σ=\( \sqrt{\frac{p(1-p)}{n}} \) into the z-score equation, then use Table A1 as usual
- Another way to approach this problem is to put it in terms of the count X (recall: x = np) and use the normal approximation to the binomial given above.
For small sample sizes (either np < 10 or n(1-p) < 10)
- Put the problems in terms of the count X and use the small sample Binomial probabilities.

Sample Means:

For the interpretations: note that words that appear in bold are key; words that are italics are “blanks” that need to be filled in based on the context of the problem.
When asked to find probabilities based on the sample mean $\bar{x}$:
For large sample sizes ($n \geq 30$) – it doesn’t matter what the exact distribution of $X$ is
- Distribution: approximately Normal($\mu$, $\sqrt{\frac{\sigma}{n}}$)
  - This Result is called the Central Limit Theorem (or CLT)
  - To find probabilities: compute the z-score by plugging $\mu=\mu$ and $\sigma=\sqrt{\frac{\sigma}{n}}$ into the z-score equation, then use Table A1 as usual
For small sample sizes, find probabilities using the same approach if $X$ comes from a Normal distribution.
Population Proportion (top of page 2 on yellow card):
Notes about confidence intervals (CI):
- An approximate 95% conservative CI for the population proportion is: $\hat{p} \pm \frac{1}{\sqrt{n}}$
  - This is a generalization of the conservative CI formula where $z*=2$
- If you want to find CI for other percentages (e.g. a 90% CI), you need to find the appropriate $z*$ using Table A1 and plug it into one of the formulas given (either conservative or general).
- The probability that the population parameter is in any given confidence interval is either 0 (meaning the parameter is not in the interval) or 1 (meaning the parameter is in the interval).
- Interpretation of the 95% Confidence Interval: We are 95% confident that the true population parameter is contained in this interval. (change % as appropriate)
- Interpretation of the 95% Confidence Level: If this procedure were repeated many times, 95% of the resulting CI would contain the true population parameter. (change % as appropriate)
Notes about sample size:
- $m$ is the margin of error, which will usually be given to you in the context of a sample size problem.
- If your answer is not a whole number, always round UP to the next whole number.

Below is some helpful information that isn’t directly shown on the formula card.
Uniform random variables:
- Problem will usually say that a variable is uniformly distributed
- Distribution: Uniform(a,b)
- To find probabilities: The uniform distribution looks like a rectangle, so probabilities are equal to the area of a rectangle (=base*height)
Quick review of graphs:
- Histogram: ONLY graph used to show the general shape of a distribution
- QQ-plot: if data points fall on an upward sloping line, then it is reasonable to assume that the data follows a normal distribution
- Side-by-side boxplots are good for comparing distributions (5-number summary)
- Bar chart: visual display of category counts (for discrete random variable)

For the interpretations: note that words that appear in bold are key; words that are italics are “blanks” that need to be filled in based on the context of the problem.
- Time series plot: Comment on stability of mean and variance; if EITHER of these are unstable, then overall the process is unstable and it may not be appropriate to treat the data a random sample.

Be careful with your language when interpreting these plots! Use words like “approximately” or “appears to be”. For example, “the data is roughly bell-shaped” or “there appears to be a trend in the mean over time”.

Basic Notation:

<table>
<thead>
<tr>
<th></th>
<th>Population Parameter</th>
<th>Sample Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>( \mu )</td>
<td>( \bar{x} )</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>( \sigma )</td>
<td>( s )</td>
</tr>
<tr>
<td>Proportion</td>
<td>( p )</td>
<td>( \hat{p} )</td>
</tr>
</tbody>
</table>

- If a problem says you have a random sample, then the ensuing information is sample information \((x, s, \text{ or } \hat{p})\). If the problem doesn’t mention a sample, then it’s most likely talking about population information \((\mu, \sigma, \text{ or } p)\).