This problem set covers material from Week 7, dates 10/24- 10/27. Unless otherwise noted, all problems are taken from the textbook. Problems can be found at the end of the corresponding chapter. "AP" stands for additional problems not found in the book.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Tuesday 10/24

- **Chapter 5**: 5, 13a and b, 18

Thursday 10/26

- **Chapter 5**: 22, 29
- **AP 1**: Suppose you randomly sample n observations $\{x_1, x_2, \ldots, x_n\}$ from a population and want to verify that the underlying population follows a certain distribution with CDF F. For example, suppose $\{x_1, x_2, \ldots, x_n\}$ represent the weights of n randomly selected newborn babies at a hospital, and that we are interested in determining whether this sample gives evidence that the weight of newborns, in general, is approximately Normally distributed (with mean weight 3.5kg and standard deviation 0.4 kg).

  One way to assess the distribution is to create a **histogram** of the sample, which divides the range of the variable into equal length intervals, counts the number of sampled values in each interval, and plots the results using rectangles with bases on each interval and with heights corresponding to the counts in each interval. A histogram of a random vector `v` can be created in R using the `hist(v)` function. If the population is approximately Normal, then the outline of the histogram should look approximately like the Normal density.

  a) Simulate a sample of 250 observations from a Normal distribution with mean of 3.5 and standard deviation of 0.4 and create a histogram of the results.
  b) Create a plot of the PDF of an $N(3.5, 0.4^2)$ variable on the interval corresponding to the observed values of your sample. *Note, there are two different ways to do this!* Compare the shape of the Normal PDF curve to the outline of the histogram. How similar do they look? Based on the shape of the histogram *alone*, would you be confident in saying the underlying population is Normally distributed?
  c) Repeat part (a), but this time with a sample of 9 observations from $N(3.5, 0.4^2)$. Based on the shape of the histogram *alone*, would you be confident in saying the underlying population is Normally distributed?

Friday 10/27

- **Chapter 5**: 37 (a-c only)
- **AP 2** (adaption of Blitzstein 5.46): Let T be the lifetime of a person (how long that person lives), with CDF $F_T(t)$ and PDF $f_T(t)$ for $t > 0$. The *hazard function* of T is defined by

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)}$$

  In common language, the *hazard* is the probability of the event (i.e. death) occurring during any given instant.

  a) Find the conditional CDF of T given that the person has survived to at least time t_0. That is, find $P(T \leq t | T \geq t_0)$ for $t \geq t_0 > 0$.

  b) Using your conditional CDF from (a), find the conditional PDF of T given that the person has survived to at least time t_0. How does this relate to the hazard function? Briefly explain why this makes sense/give an interpretation to the hazard function.

  c) Show that an Exponential(λ) random variable has a constant hazard function.

- **AP 3**: Let $Y \sim$ Exponential(λ). Show that $\mathbb{E}[Y] = \frac{1}{\lambda}$. Integration by parts will be helpful: $\int_a^b u dv = uv|_a^b - \int_a^b v du$

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.