Instructions: This homework assignment covers **Chapter 5** of the course notes. On each part, there is opportunity for partial credit, so show all of your work and explain all of your reasoning. No work/no explanation means no credit even if your answer is correct. If you use R to answer any part or to check your work, please include all code and output as attachments. Do not just write out the code you used.

1. In a recent study, biomedical engineers and nephrologists modeled the time to death (T, measured in months) for dialysis patients with chronic kidney disease using a Weibull distribution with shape parameter $\beta = 1.4$ and scale parameter $\eta = 17.5$.

(a) Use R to graph the probability density function $f_T(t)$.
(b) Find the mean and median time to death. Why is the mean larger?
(c) Find the survivor function $S_T(t)$ and calculate $S_T(24)$. What does $S_T(24)$ represent?
(d) Ninety-five percent of the patients will die before what time?
(e) Is the hazard function $h_T(t)$ increasing, decreasing, or constant? Explain what this implies about the population of dialysis patients with chronic kidney disease.

2. The data below are taken from Table 1 in Xu et al. (2003, *Applied Soft Computing*), who describe a reliability study on turbochargers in diesel engines. These are failure time data for $n = 40$ turbochargers. The failure time T is measured in 1000s of hours.

1.6	2.0	2.6	3.0	3.5	3.9	4.5	4.6	4.8	5.0
5.1	5.3	5.4	5.6	5.8	6.0	6.0	6.1	6.3	6.5
6.5	6.7	7.0	7.1	7.3	7.3	7.3	7.7	7.7	7.8
7.9	8.0	8.1	8.3	8.4	8.4	8.5	8.7	8.8	9.0

(a) Assume a Weibull distribution for T. Under this assumption, use R to find the maximum likelihood estimates $\widehat{\beta}$ and $\widehat{\eta}$.
(b) What does the (estimated) hazard function look like? Explain what this implies about the population of turbo chargers.
(c) Does the Weibull distribution seem reasonable for these data? Construct a Weibull qq plot and make your assessment.

3. A continuous distribution regarded as a "competitor" to the Weibull distribution as a model for lifetime random variables is the **lognormal distribution**, whose pdf is given by

$$f_T(t) = \begin{cases} \dfrac{1}{t\sqrt{2\pi\sigma^2}} \exp\left[-\dfrac{1}{2\sigma^2} (\ln t - \mu)^2\right], & t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The parameters μ and σ^2 are not the mean and variance in this distribution like they are in the normal distribution.

LOGNORMAL R CODE: Suppose $T \sim \text{lognormal}(\mu, \sigma^2)$.

$F_T(t) = P(T \leq t)$	ϕ_p
`plnorm(t,`μ`,`σ`)`	`qlnorm(p,`μ`,`σ`)`

Suppose the lifetime T (measured in hours) of a semiconductor laser has a lognormal distribution with $\mu = 5.5$ and $\sigma^2 = 0.5$.

(a) Use this code to graph the probability density function (pdf) of T:

```
# Plot PDF
t = seq(0,2000,1)
pdf = dlnorm(t,5.5,sqrt(0.5))
plot(t,pdf,type="l",xlab="t",ylab=expression(f[T](t)),cex.lab=1.25)
abline(h=0)
```

(b) The mean of a lognormal random variable is not μ. Instead, it is

$$E(T) = e^{\mu+\sigma^2/2}.$$

Calculate the mean semiconductor laser lifetime.
(c) What is the probability a semiconductor laser lifetime is larger than 1,000 hours? *Hint:* Use the `plnorm` function in R. Note that this function uses σ; not $\sigma^2 = 0.5$.
(d) Ten percent of the semiconductor lasers will fail before what time? *Hint:* Use the `qlnorm` function in R. Note that this function uses σ; not $\sigma^2 = 0.5$.
(e) The hazard function for a lognormal random variable T is

$$h_T(t) = \frac{\frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\,(\ln t - \mu)^2\right]}{1 - F_Z\left(\frac{\ln t - \mu}{\sigma}\right)},$$

for $t > 0$, where $F_Z(\cdot)$ denotes the $\mathcal{N}(0,1)$ cdf. Comparing this to the Weibull hazard function, why do you think engineers generally prefer the Weibull distribution over the lognormal distribution when modeling lifetimes?
(f) Of course, just because the Weibull distribution is used more often doesn't mean that it is always a good choice. Suppose you (incorrectly) assumed the semiconductor laser lifetime distribution was Weibull when, in reality, it is lognormal. What are potential consequences of using an incorrect model?