Note: This homework assignment covers Chapter 5.

Disclaimer: If you use R, include all R code and output as attachments. Do not just "write in" the R code you used. Also, don't just write the answer and say this is what R gave you. If my grader can't see how you got an answer, it is wrong. I want to see your code and your answers accompanying your code (like in the notes).

1. The time to failure (T, measured in hours) for a piece of electronic equipment used in the manufacture of Blue Ray players has the following probability density function (pdf):

$$f_T(t) = \begin{cases} \frac{1}{2000} e^{-t/2000}, & t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Note that this is the pdf of an exponential random variable T with $\lambda = 1/2000$.
(a) What is the probability that this piece of equipment will fail before 4000 hours?
(b) Find the median of this distribution. Interpret, in words, what the median represents.
(c) If you graphed the hazard function of T as a function of t (time), what would it look like? Explain why.
(d) Graph the survivor function $S_T(t)$. Interpret what $S_T(3000)$ means in words.

2. A shock absorber is a suspension component that controls the up-and-down motion of a vehicle's wheels. The following data are $n = 38$ distances (in km) to failure for a certain brand of shock absorber under "extreme" driving conditions.

6700	6950	7820	9120	9660	9820	11310	11690	11850	11880
12140	12200	12870	13150	13330	13470	14040	14300	17520	17540
17890	18450	18960	18980	19410	20100	20100	20150	20320	20900
22700	23490	26510	27410	27490	27890	28100	30050		

(a) Under a Weibull assumption for

$$T = \text{ distance (in km) to failure}$$

calculate the maximum likelihood estimates $\widehat{\beta}$ and $\widehat{\eta}$ for the data above. Use these values (along with the Weibull assumption) to answer the following questions.
(b) Calculate $P(T > 15000)$. Interpret what this probability means in words.
(c) Find the 25th percentile of the distance to failure distribution. Interpret in words.
(d) Plot the estimated hazard function for T. Explain, in plain English, what information this graph reveals.
(d) Does the Weibull model seem reasonable for these data? Construct a Weibull qq plot. Interpret this plot. In the light of your answer here, how do you feel about the accuracy of your answers to parts (b), (c), and (d)?

3. A probability distribution well known as a "competitor" to the Weibull distribution for modeling lifetime data is the **lognormal distribution**, whose pdf is given by

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

The parameters μ and σ^2 are not the mean and variance in this distribution (like they are in the normal distribution). Suppose that the lifetime T (measured in hours) of a semiconductor laser has a lognormal distribution with $\mu = 10$ and $\sigma^2 = 2.25$.
(a) Use this code to graph the probability density function (pdf) of T:

```
# Plot PDF
t = seq(0,300000,1)
pdf = dlnorm(t,10,sqrt(2.25))
plot(t,pdf,type="l",xlab="t",ylab="f(t)")
abline(h=0)
```

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

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

(b) What is the probability that the lifetime exceeds 50,000 hours? Go to your graph for the pdf and show what this probability represents.
(c) Find the proportion of lifetimes between 100,000 hours and 200,000 hours.
(d) Ten percent of lifetimes will greater than what value? Show where this corresponding percentile falls on the pdf.
(e) The hazard function for $T \sim \text{lognormal}(\mu, \sigma^2)$ is given by

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

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

4. The time to failure (in hours) of a bearing used in a mechanical shaft is under investigation. A sample of $n = 84$ bearings produced the observations in Table 1 (next page). Load the `MASS` and `car` functions in R. Enter the data in Table 1 (carefully!) and call the data vector `bearing`.

(a) Fit a Weibull distribution to the data. Use the code:

```
# Fit Weibull model
fitdistr(bearing,densfun="weibull")
```

(b) Construct a Weibull qq plot for the data using the code

```
qqPlot(bearing,distribution="weibull",shape=beta.hat,scale=eta.hat,pch=16)
```

27135.9	3293.5	16380.9	3406.6	12455.7	1014.4	4627.3
22800.0	425.5	30859.3	202.4	3508.5	5475.3	6924.4
534.9	288.6	1363.4	9200.5	154.3	325.2	4197.5
5592.2	682.1	172.3	25287.2	125.0	1753.7	1010.5
5118.9	2217.6	175.8	3142.7	5094.3	33154.8	427.5
6142.1	15180.4	971.4	103.2	2691.4	2406.8	3961.8
1814.8	1074.0	812.8	21022.6	4548.1	9877.8	707.9
1903.9	2293.7	2581.4	31597.1	25994.2	3661.3	2004.0
238.5	2164.7	22304.2	15.9	157.8	17673.4	143.0
672.2	2671.0	417.8	5421.3	290.9	17286.1	4263.6
8492.6	8885.1	16947.8	29890.5	4102.9	11009.1	15.3
5663.3	41.4	2002.8	1329.6	29821.9	139.0	130.6

Table 1: Bearing data for Problem 4.

(c) Now, repeat for the lognormal distribution:

```
# Fit lognormal model
fitdistr(bearing,densfun="log-normal")
qqPlot(bearing,distribution="lnorm",pch=16)
```

(d) Which distribution appears to be a better fit: the Weibull or the lognormal?
(e) If you want to see what a really bad model fit looks like, type

```
qqPlot(bearing,distribution="norm",pch=16)
```

This is the qq plot under a normal (Gaussian) assumption. Type `hist(bearing)` to see a histogram of the bearing data; you can see it does not resemble a normal distribution whatsoever.