Question 1.
(a) There are

$$\binom{66}{5} = \frac{66!}{5!\ 61!} = \frac{66(65)(64)(63)(62)}{120} = 8,936,928 \text{ possible samples.}$$

(b) Y follows a **hypergeometric** distribution. There are only 6 possible values of Y: 0, 1, 2, 3, 4, and 5. We can list out the values of Y. This makes Y discrete.
(c) Use hypergeometric pmf:

$$P(Y = 4) = \frac{\binom{36}{4}\binom{30}{1}}{\binom{66}{5}}$$

We have

$$\begin{aligned} \binom{36}{4} &= \frac{36!}{4!\ 32!} = \frac{36(35)(34)(33)}{24} = 58905 \\ \binom{30}{1} &= 30. \end{aligned}$$

Therefore,

$$P(Y = 4) = \frac{58905(30)}{8936928} \approx 0.198.$$

Question 2.
(a) If the circuits are laid out in a series, then the product will function only if **all** the circuits are functioning. If the circuits are laid out in parallel, then the product will function as long as **at least one** circuit is functioning.

The easiest way to derive each formula is to let

$$Y = \text{ number of functioning circuits}$$

and recognize that $Y \sim b(n, p)$, provided that the Bernoulli trial assumptions hold (independent circuits; same probability p of functioning for each; each circuit functions/not). For a series system,

$$r_n = P(Y = n) = \binom{n}{n} p^n (1-p)^0 = p^n.$$

For a parallel structure,

$$\begin{aligned} r_n = P(Y \geq 1) &= 1 - P(Y = 0) \\ &= 1 - \binom{n}{0} p^0 (1-p)^n \\ &= 1 - (1-p)^n. \end{aligned}$$

(b) We are left to solve

$$1 - (1-p)^6 = 0.9999 \quad \text{for } p.$$

$$\begin{aligned} (1-p)^6 = 0.0001 \quad &\Longrightarrow \quad 1 - p = (0.0001)^{1/6} \\ &\Longrightarrow \quad p = 1 - (0.0001)^{1/6} \approx 0.785. \end{aligned}$$

Question 3.

(a) The median $\phi_{0.5}$ solves the following equation:

$$\int_5^{\phi_{0.5}} \frac{3}{8}(7-y)^2 dy = 0.5.$$

Interpretation: 50 percent of the water samples will have pH below 5.413; 50 percent of the water samples will have pH above 5.413.

(b) Think of each water sample as a "trial." Think of the event that a water sample's pH exceeds 6.0 as a "success." The number of water samples, say X, required to find the first "success" follows a geometric distribution with

$$p = P(Y \geq 6.0) = \int_6^7 \frac{3}{8}(7-y)^2 dy.$$

To do this integral cleanly, I let $u = 7 - y$ so that $du = -dy$. Therefore,

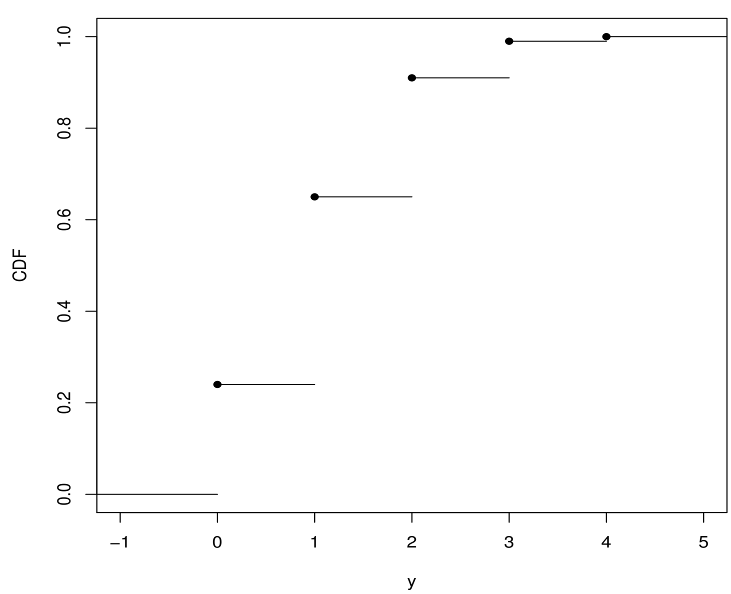
$$\begin{aligned} \int_6^7 \frac{3}{8}(7-y)^2 dy &= \frac{3}{8}\int_0^1 u^2 du \\ &= \frac{3}{8}\left(\frac{u^3}{3}\bigg|_0^1\right) = \frac{1}{8}. \end{aligned}$$

Therefore, the probability that we will need 1 or 2 water samples is

$$P(X=1) + P(X=2) = (1-p)^0 p + (1-p)^1 p = \frac{1}{8} + \frac{7}{8}\left(\frac{1}{8}\right) \approx 0.234.$$

Question 4.

(a) Because I am typing out the solutions, I will construct the CDF in R.



(b) First calculate $E(Y)$ and $E(Y^2)$:

$$\begin{aligned} E(Y) &= 0(0.24) + 1(0.41) + 2(0.26) + 3(0.08) + 4(0.01) \approx 1.21 \\ E(Y^2) &= 0^2(0.24) + 1^2(0.41) + 2^2(0.26) + 3^2(0.08) + 4^2(0.01) \approx 2.33. \end{aligned}$$

Therefore, the expected weekly attendance score is

$$\begin{aligned} E(W) &= E(6 - 0.5Y - 0.25Y^2) \\ &= 6 - 0.5E(Y) - 0.25E(Y^2) \\ &\approx 6 - 0.5(1.21) - 0.25(2.33) \approx 4.813. \end{aligned}$$

Question 5.
(a) These are the values of β and η that maximize the likelihood function. Or, these are the values of β and η that most closely agree with the observed data.
(b) With $\widehat{\beta} = 0.66$ and $\widehat{\eta} = 5115.83$, we calculate

$$\begin{aligned} P(T < 15000) &= F_T(15000) \\ &= 1 - e^{-\left(\frac{15000}{5115.83}\right)^{0.66}} \approx 0.869. \end{aligned}$$

(c) The survivor function $S_T(t)$ is simply $1 - F_T(t)$. Therefore, for $t > 0$,

$$h_T(t) = \frac{f_T(t)}{S_T(t)} = \frac{\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-(t/\eta)^\beta}}{e^{-(t/\eta)^\beta}} = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}.$$

Plugging in the maximum likelihood estimates, we have

$$h_T(t) = \frac{0.66}{5115.83}\left(\frac{t}{5115.83}\right)^{0.66-1}.$$

Note that $\widehat{\beta} = 0.66 < 1$. This means the estimated hazard function is a decreasing function of time. This suggests the rate of failure decreases over time (i.e., the population of bearings is getting stronger over time).
(d) **My first reaction:** There is a pretty noticeable departure from linearity in this plot. This might make me question the Weibull assumption for bearing lifetimes. This, in turn, would make me question the trustworthiness of my answers in parts (b) and (c).

On the other hand, all of the bearing data do fall within the "bands of uncertainty," so perhaps the departure we are seeing is not enough to discount the Weibull model from being reasonable.

Only the Oracle knows for sure!