STAT 641 - Partial Solutions - EXAM I

Part I. (40 points)

1. **A:** N is the number of batches out of the 30 batches having $M > 2500$. This is a sequence of iid Bernoulli trials. Thus, the binomial distribution with n=30 and $p = P[M > 2500]$ would be a reasonable model for N.

2. **C:** This study is a stratified simple random sample with Strata being the ten Relative Humidity Levels and Sampling Unit is a Power Plant.

3. **D:** The value of E corresponding to $U = 0.43$ is obtained from the Poisson distribution with $\lambda = (5)(1) = 5$. Using the Poisson tables with $\lambda = 5$ the following values are obtained from the cdf, F(y), for the Poisson distribution or are obtained by summing the terms in the Poisson pmf $f(y) = \lambda^y e^{-\lambda}/y!$ for $y = 0, 1, 2, \ldots$ with $\lambda = 5$.

y	0	1	2	3	4
f(y)	.007	.033	.085	.140	.175
F(y)	.007	.040	.125	.265	.440

Therefore, the randomly generated value of E is 4, the smallest value of y such that $F(y) = P[Y \leq y] \geq 0.43$.

4. **E:** All of the statements are true.

5. **B:** The contribution of $Y_{15} = 5.4$ to the kernel density estimate of $f(4)$ is obtained from

$$\frac{1}{(30)(2)} K\left(\frac{4-5.4}{2}\right) = \frac{1}{(30)(2)}\left(1 - \left|\frac{4-5.4}{2}\right|\right) = .005$$

6. **A:** As the bandwidth, h, increases $\hat{f}(y)$ incorporates more of the data values into the estimate and hence smoothes the estimate.

7. **A:** $\widehat{Q}(u) = Y_{(n+1)u}$ along with n=20, yields

$\widehat{Q}(.8) = Y_{(20+1).8} = Y_{16.8} = Y_{16} + .8(Y_{17} - Y_{16}) = 26.6 + .8(27.1 - 26.6) = 27.0$

8. **D:** MAD is preferred to standard deviation as a measure of population dispersion when the population distribution has a heavy-tailed distribution.

9. **E:** All of the statements are false.

10. **B:** The recorded value of the amount of retained water for the censored ticks will be greater than their value at 90 days thus the type of censoring is Left censored

Part II. (60 points)

1. (9 points each) The survival function is $S(v) = 1 - F(v) = e^{-(v/\alpha)^\gamma}$ and the quantile function is

   $Q(u) = \alpha[-log(1-u)]^{1/\gamma}$.

   From the SAS output, the estimators of the Weibull parameters which incorporate censoring are given by $\hat{\alpha} = 118.3669$ and $\hat{\gamma} = 0.694$. Therefore,

   a. $\hat{\tilde{\mu}} = \hat{Q}(.5) = \hat{\alpha}[-log(1-u)]^{1/\hat{\gamma}} = (118.3669)(-log(1-.5))^{1/.694} = 69.80$

   b. $\hat{S}(80) = e^{-(80/118.3669)^{.694}} = .467$:

   c. V_W is that value of V such that $P[V \geq V_W] = .9$, that is, $P[V < V_W] = .1$, which yields $V_W = Q_V(.1)$. Therefore, the MLE estimate is

      $\hat{V}_W = \hat{Q}_V(.1) = \hat{\alpha}[-log(1-.1)]^{1/\hat{\gamma}} = (118.3669)(-log(1-.1))^{1/.694} = 4.623$

2. (9 points each) Using the Kaplan-Meirer Product Limit estimator produced by SAS, the three estimates are

   a. $\hat{Q}(u) = inf\left\{v : \hat{S}(v) \leq 1-u\right\} \Rightarrow \hat{Q}(.5) = inf\left\{v : \hat{S}(v) \leq .5\right\} = 55.3$

   b. $\hat{S}(80) = .42 + \frac{(80-83.6)(.44-.42)}{.77-83.6} = 0.431$

   c. $V_W = Q(.1) \quad \Rightarrow \quad \hat{V}_W = \hat{Q}(.1) = inf\left\{v : \hat{S}(v) \leq 1-.1\right\} = 6.4$

3. (6 points) Ignoring that some of the data values are right censored, the Weibull model appears to provide a reasonable model for the failure voltages for the following reasons:

   i. The estimated pdf for V is a highly right skewed pdf

   ii. The plotted points in the Weibull reference plot are reasonable close to a straight line.

   - However, ignoring the censoring could result in a very misleading conclusion. Thus the reference distribution plot should be modified to take into account the censoring. Using the estimates which take into account the censoring we have the following results:

     The Weibull based estimates of the median and V_W are not very close to the distribution-free estimates obtained from the Product Limit Estimator of S(t). Likewise, the Kaplan-Meirer PL estimate of V_W is about 50% larger than the Weibull based estimator. Therefore, I would tentatively conclude that the Weibull model is not a good fit to the voltage data.

     Median: 69.8 vs 55.3 S(80): .467 vs .43 V_W: 4.623 vs 6.4

Exam 1 Scores for STAT 641 - Fall 2019

Min = 37, Q(.25) = 61.3, Q(.5) = 83.5, Mean = 75.7, Q(.75) = 91, Max = 100