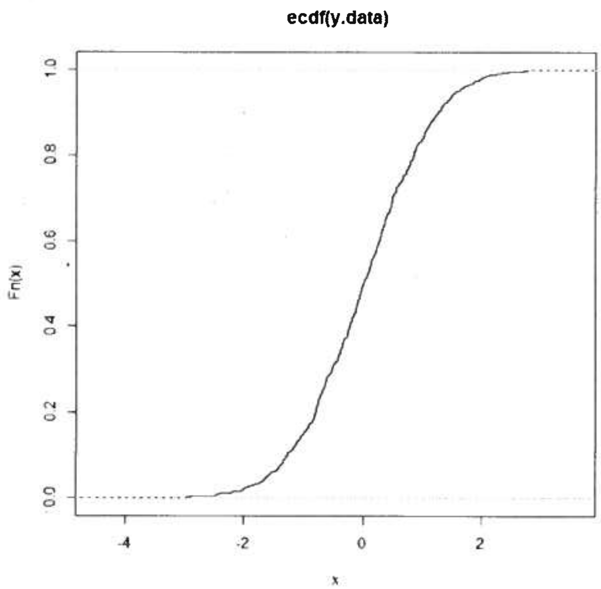
Version B Solutions

Question	Marks
1	6
2	8
3	11
4	8
Total	33

Aids: Only question pages will be marked. Final numerical answers must be rounded to exactly 3 decimal places, unless the answers are exact to less than 3 decimal places. For intermediate steps, please be sure to keep all decimal places to avoid round off errors.

Question 1. Multiple Choice Questions – 1 mark for each question. Please clearly circle your answer.

a. An ecdf plot of sample data labeled y.data is given below:



Which model would be the best choice for this sample data, based on the above ecdf plot?

A. Normal(0, 1) B. Poisson(01) C. Uniform(0, 1) D. Exponential(1)

b. Suppose it is reasonable to model observed data y using an Poisson(θ) distribution. Let $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$. Which of the following functions in R should be used to calculate the relative likelihood function for θ?

A. `PoisRLF <- function(x) {exp(y*log(x/thetahat)+(n-y)*log((1-x)/(1-thetahat)))}`

B. `PoisRLF<-function(x) {exp(n*thetahat*log(x/thetahat)+n*(thetahat-x))}`

C. `PoisRLF<-function(x) {(thetahat/x)^n*exp(n*(thetahat/x)-1)}`

D. `PoisRLF<-function(x) {(thetahat/x)^n*exp(n*(1-thetahat/x))}`

E. More than one of these.

c. Suppose $Y_i \sim G(\mu, 5)$, for i = 1, 2, ..., n; Independent. What minimum sample size is needed so that $P(|\bar{Y} - \mu| \leq 1) \geq 0.99.$ → z-score = 2.576

A. 166 B. 25 C. 34 D. 13

$P\left(\frac{|\bar{Y}-\mu|}{\sigma/\sqrt{n}} \leq \frac{1\sqrt{n}}{5}\right) \geq 0.99$ → Set $\frac{\sqrt{n}}{5} = 2.576 \Rightarrow n = (2.576 \times 5)^2$

$\Rightarrow n = 165.89 \rightarrow n = 166$

d. Suppose $Y_i \sim$ Poisson(49), i = 1, 2, ..., n; Independent. A random sample of 49 observations, $y_1, y_2, \ldots, y_{49}$ is drawn from this population. According to the Central Limit Theorem:

$E(Y_i) = 49$, $Var(Y_i) = 49$.

$E(\bar{Y}) = \mu = 49$

$Var(\bar{Y}) = \frac{\sigma^2}{n} = \frac{49}{49} = 1 \Rightarrow$ Std. dev. $(\bar{Y}) = 1$

A. the sampling distribution of $\bar{Y}$ is G(49, 1) exactly

B. the sampling distribution of $\bar{Y}$ is G(49, 7) exactly

C. the sampling distribution of $\bar{Y}$ is G(49, 1), approximately

D. the sampling distribution of $\bar{Y}$ is G(49, 7), approximately

E. The Central Limit Theorem doesn't apply here, as the parent population is Poisson

e. Suppose that a random sample of 64 U. Waterloo students was selected. The students were asked if they thought tuition was too high. Forty-eight said "yes". An approximate 90% confidence interval for θ, the true proportion of U. Waterloo students who think that tuition is too high, is given by:

A. (0.644, 0.856) B. (0.661, 0.839) C. (0.696, 0.804) D. (0.745, 0.755)

$\hat{\theta} = \frac{48}{64} = 0.75$. An approx. 90% CI for θ is:

$\hat{\theta} \pm 1.645\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \rightarrow 0.75 \pm 1.645\sqrt{\frac{0.75(0.25)}{64}}$

0.75 ± 0.089

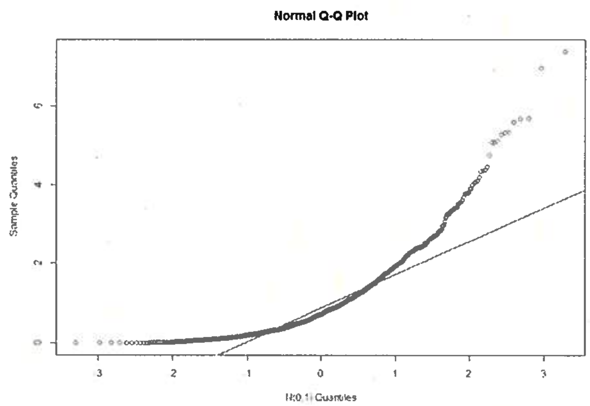
f. Suppose that weights of male University students are Normally distributed with unknown mean, μ kg, and a known variance of 100. A random sample of 16 weights yielded a sample mean of 83. Based on this information, a 95% confidence interval for the mean weight of male University students is:

A. (78.1, 87.9) B. (78.89, 87.11) C. (34, 132) D. (80.5, 85.5)

A 95% CI for μ is: $\bar{y} \pm 1.96\frac{\sigma}{\sqrt{n}} \rightarrow 83 \pm 1.96\left(\frac{10}{\sqrt{16}}\right) \rightarrow 83 \pm 4.9$

Question 2. True / False - 1 mark for each question. Record your answer in the blank. No need to show your work here.

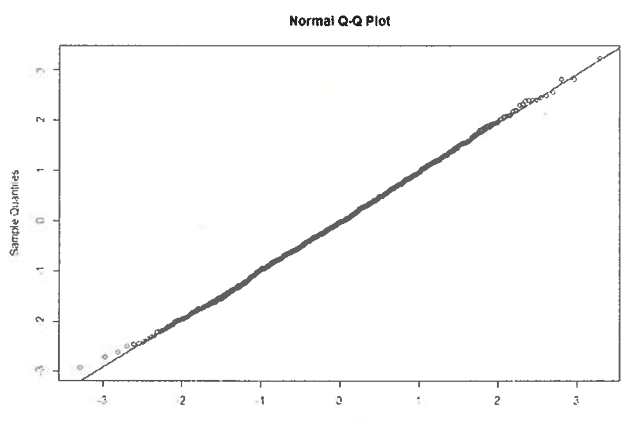
a. Consider the following Normal QQ Plot:



The sample data suggest that a model that is positively skewed, like the Exponential would be more appropriate.

-distinct U-shape → positively skewed. T

b. Consider the following Normal QQ Plot:



The sample data suggest that a model that has thinner tails than the Normal, like the Uniform would be more appropriate.

-A Normal model should be used here F

c. Suppose Y~ Exponential(θ). The following pivotal quantity, $\frac{\bar{Y}-E(\bar{Y})}{std.\ dev(\bar{Y})}$ has an exact G(0, 1) distribution.

$Y \sim Exp(\theta)$. $\tilde{\theta} = \bar{Y}$ $\quad \frac{\bar{Y}-E(\bar{Y})}{std.dev(\bar{Y})} \overset{\cdot}{\sim} G(0,1)$ (by the CLT). F

d. A 90% confidence interval for μ was constructed based on a random sample of n = 100 observations drawn from a Normal parent population, with a known population standard deviation, σ. If a random sample four times as large had been taken (i.e, n = 400), the resulting 90% confidence interval would be half as wide as the original one.

Four times as many observations will result in a CI half as wide. The value n is under a square root. T

e. A 90% confidence interval for μ constructed using a sample of 80 observations drawn from a Normal parent population, with a known population standard deviation, σ. A corresponding 95% confidence interval for μ based on the same sample of observations would be wider. T

All else being equal, increasing the level of confidence will widen the CI.

f. A 20% likelihood interval was constructed, assuming a Poisson model was appropriate. A random sample of 36 observations resulted in a sample mean of 4.5, so the mle of θ was $\hat{\theta}$=4.5. A 10% likelihood interval based on the same sample of observations would be wider than the 20% likelihood interval.

A 10% likelihood interval will be wider than a 20% likelihood interval, all else being equal. T

g. If the attributes of interest in the sample differ from the attributes of interest in the target population, the difference is called sample error.

Sample error occurs when the attributes of interest in the study pop^n differ from the attributes in the sample. F

h. If the attributes of interest in the sample differ from the attributes of interest in the study population, the difference is called study error.

-This is called sample error. (see above). F

Study error occurs when the attributes in the target pop^n differ from the attributes in the study pop^n.

Question 3. Written Answer Question – Show ALL necessary work for full credit.

[11] Suppose that three identical six-sided dice, each with faces marked 1 to 6 are rolled 200 times. After each roll, the number of dice showing an even-number on the upper face is recorded. The results have been summarized in the table below:

# of Even Faces	Observed	Expected
0	26	23.5
1	70	73.5
2	76	76.5
3	28	26.5

$\sum$ sum = 200 ✓

a. ***[2]*** Explain why a Binomial model would be appropriate.

- Two outcomes per trial.
- Independent trials (Bernoulli trials)
- P(success) = P(rolling an even #) stays constant from trial to trial.

b. ***[2]*** Show that the mle of θ, the probability of rolling an even number, is $\hat{\theta}$= 0.51. **Note:** If you cannot show this, move on to the next part of the question, and use $\hat{\theta}$= 0.51.

Avg. # of even faces each game $= \dfrac{0(26)+1(70)+2(76)+3(28)}{200} = 1.53$

There are 3 dice rolled in each game $\Rightarrow$ $\hat{\theta} = \dfrac{1.53}{3} = 0.51$ //

c. ***[5]*** Assuming a Binomial model, with the estimate from part b., calculate the expected frequencies and fill in the table above. You can give expected values to 1 decimal place in the table. Show all calculations!

$e_0 = 200 \times \hat{P}(\text{no even \#'s}) = 200\left[\binom{3}{0} 0.49^3\right] = 23.5$

$e_1 = 200 \times \hat{P}(\text{one even \#}) = 200\left[\binom{3}{1} 0.51^1 (0.49)^2\right] = 73.5$

$e_2 = 200 \times \hat{P}(\text{two even \#'s}) = 200\left[\binom{3}{2} 0.51^2 (0.49)^1\right] = 76.5$

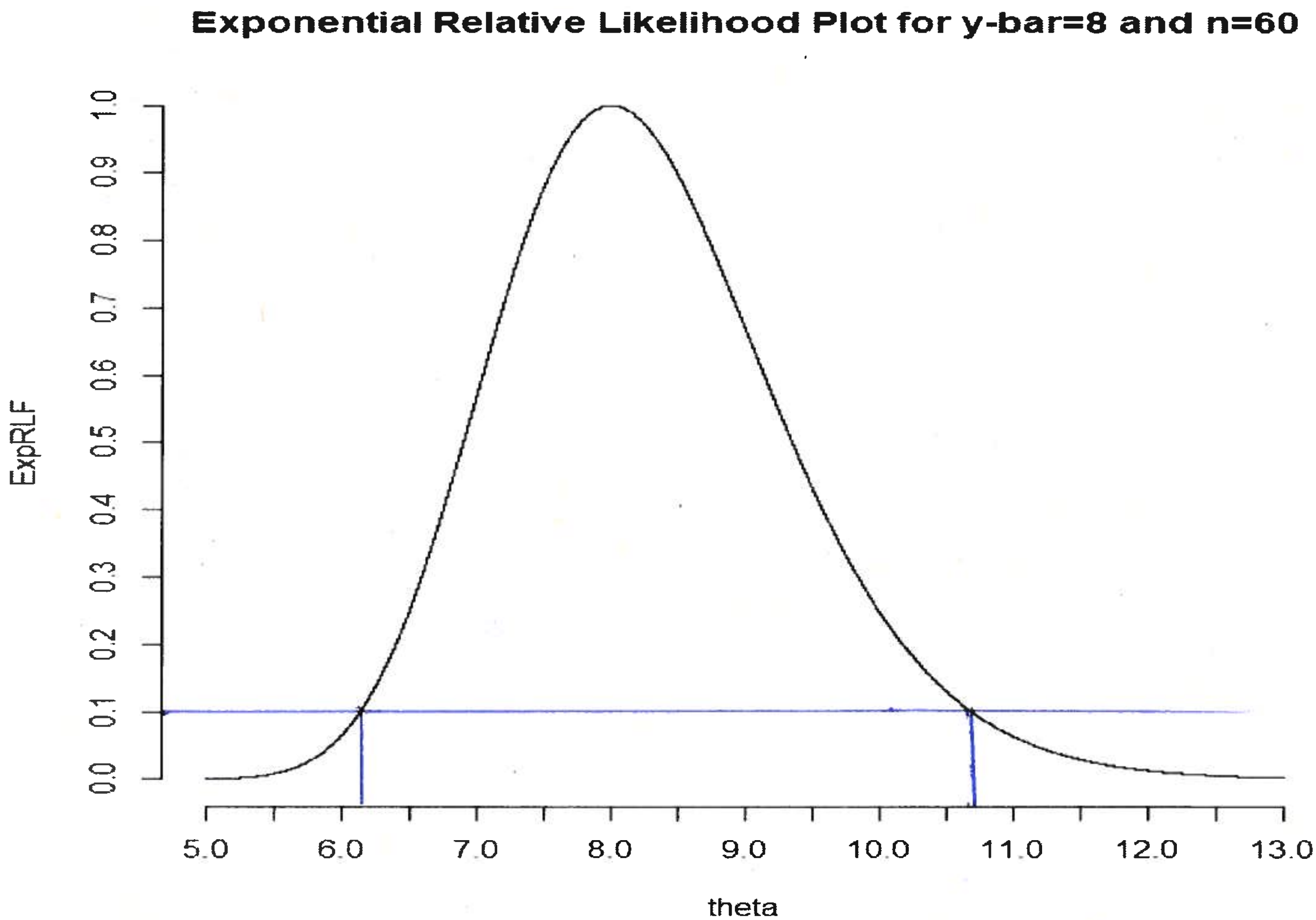
$e_3 = 200 \times \hat{P}(\text{three even \#'s}) = 200\left[\binom{3}{3} 0.51^3\right] = 26.5$

d. ***[2]*** How well does the data fit the model? You don't have to do any formal analysis here, simply compare the observed an expected frequencies and comment.

This looks like a good fit. The expected values are close to the observed values.

Question 4. Short Answer. Record your answer to each question in the blank.

[8] Given the following relative likelihood plot for a random sample of 60 independent Exponential(θ) observations, $y_1, y_2, \ldots, y_{60}$, with a sample mean of 8.



Record your answers to the questions below in the blanks provided:

a. What is the maximum likelihood estimate of θ? $\hat{\theta} = \bar{y} = 8.$

b. Using a ruler (or straight edge) and pen / pencil, determine the 10% relative likelihood interval for θ. Please express your answers to one decimal place here.

(6.1, 10.7)

*** *This question is continued on the next page.* ***

c. Answer Yes or No. Is 5.5 a plausible value of θ, in light of the observed data? No

5.5 lies outside of the 10% likelihood interval.

d. Assuming $Y_1, Y_2, \ldots, Y_{60}$ are independent Exponential(θ) random variables, using the mle of θ in part a., determine the maximum likelihood estimate for the probability that Y is greater than 8, to 3 decimal places. **Note:** If $Y \sim \text{Exp}(\theta)$, $F(y) = 1 - e^{-y/\theta}$.

$e^{-1} = 0.368$

$$\hat{P}(Y>8) = 1 - \hat{P}(Y \le 8) = e^{-8/8} = e^{-1} = 0.368$$

e. Now suppose that a random sample of size 100 was obtained from the same population, so we will assume an Exponential model here also. The sample mean was calculated to be 8 again. Would the corresponding 10% relative likelihood interval, based on the above information be wider, narrower, or the same size when compared to the interval in part b.? Narrower.

All else being equal, increasing the sample size results in a narrower interval. More informatio

f. Using your estimate in part a., and the appropriate approximate pivotal quantity, please construct an approximate 95% confidence interval for θ. Express your answers below to 3 decimal places.

(5.976, 10.024)

$$\frac{\bar{Y} - \theta}{\tilde{\theta}/\sqrt{n}} \sim G(0,1) \text{ (by the CLT) for large } n.$$

So, an approx. 95% CI for θ is:

$$\bar{y} \pm 1.96\left(\frac{\bar{y}}{\sqrt{n}}\right) \rightarrow 8 \pm 1.96\left(\frac{8}{\sqrt{60}}\right) \rightarrow 8 \pm 2.024$$