STATISTICS 641 - ASSIGNMENT 5

DUE DATE: Noon (CDT), MONDAY, OCTOBER 25, 2021

Name ____

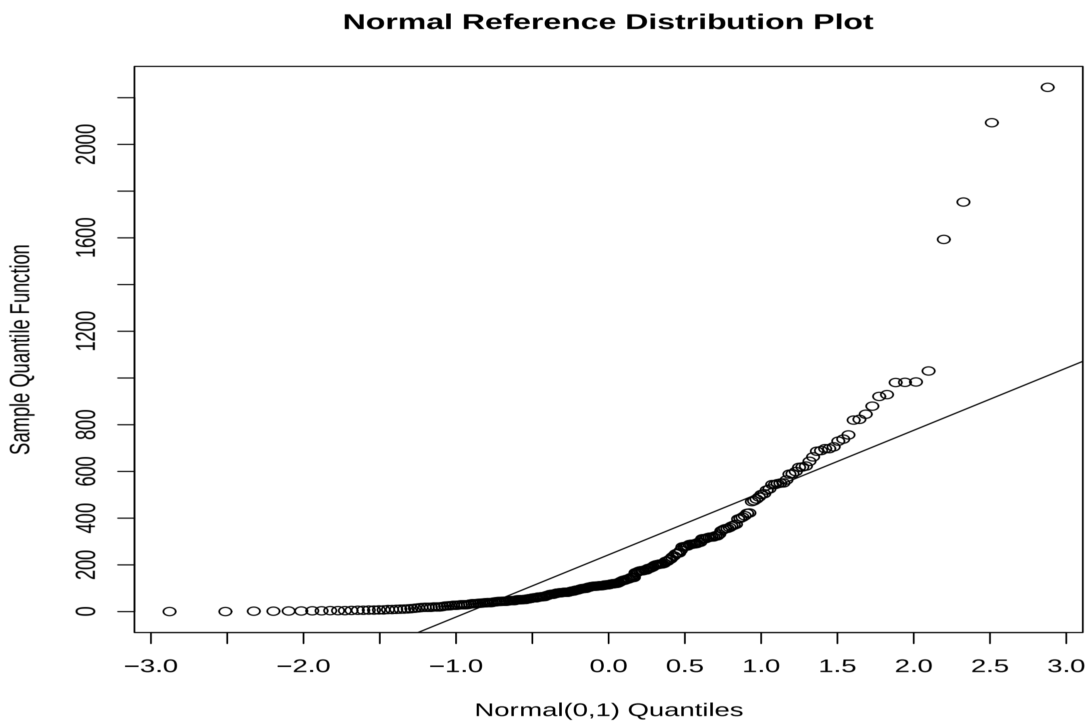
Email Address ____

Please TYPE your name and email address. Often we have difficulty in reading the handwritten names and email addresses. Make this cover sheet the first page of your Solutions.

STATISTICS 641 - ASSIGNMENT 5 - Due Noon. (CDT) Monday - 10/25/2021

- Read Handouts 8 & 9 & 10
- Supplemental Reading: Chapter 1, 6, 14, Sections 4.6, 5.3-5.6, and 7.2-7.4 in Devore book

P1. (5 points) 250 iid observations $Y_1, Y_2, \ldots, Y_{250}$ from a process yield the following normal probability plot.



- Using the above plot, describe the process distribution relative to a normal distribution with respect to tail weight and symmetry.

P2. (10 points) Nylon bars were tested for brittleness. Each of 500 bars was molded under similar conditions and was tested by placing a specified stress at 5 locations on the bar. Assuming that each bar has uniform composition, the number of breaks on a given bar should be binomially distributed with an unknown probability p of breaking. The following table summarizes the outcome of the experiment:

Breaks/Bar	0	1	2	3	4	5	total
Frequency	140	197	115	41	5	2	500

- Use a GOF test to evaluate whether the data appears to be from a binomial model.

P3. (20 points) A random sample of 500 data values are selected from four separate processes having cdf's, F_1, F_2, F_3, F_4. The plot of the sample quantile versus a standard normal quantile for each of the four samples is given below. For each of these plots, **SELECT ONE** of the following distributions to describe the pdf which generated the data. Hint: make sure to take into consideration the size of values associated with each distribution.

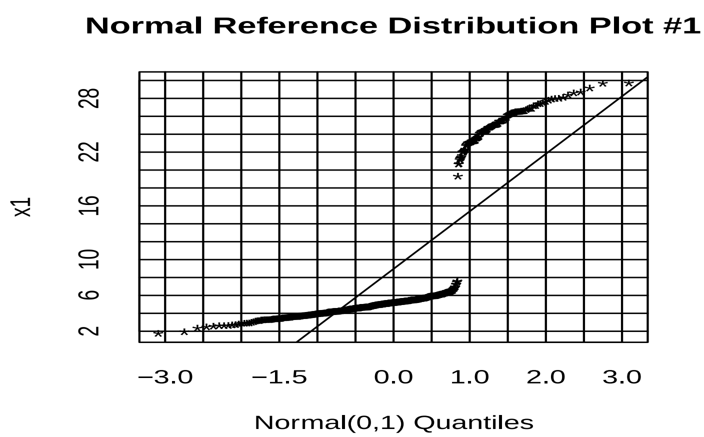
(A) Cauchy($\theta_1 = .35, \theta_2 = 1$)
(B) t with df=35
(C) Logistic($\theta_1 = .35, \theta_2 = 1$)
(D) Beta($\alpha = .1, \beta = .7$)
(E) Uniform(0,.7)
(F) Normal($\mu = .35, \sigma = 1$)
(G) Exponential($\beta = 100$)
(H) Weibull($\gamma = 1.2, \alpha = 25$)
(I) Gamma($\alpha = 2, \beta = 25$)
(J) Mixture of 90% Normal$(10, 1)$ & 10% Normal$(30, (3)^2)$
(K) Mixture of 80% Normal$(5, 1)$ & 20% Normal$(25, (3)^2)$
(L) Normal($\mu = 20, \sigma = 1$)

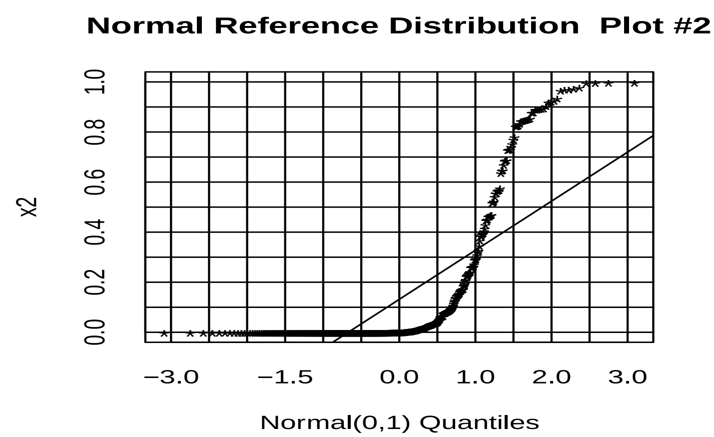
Plot 1 ___________

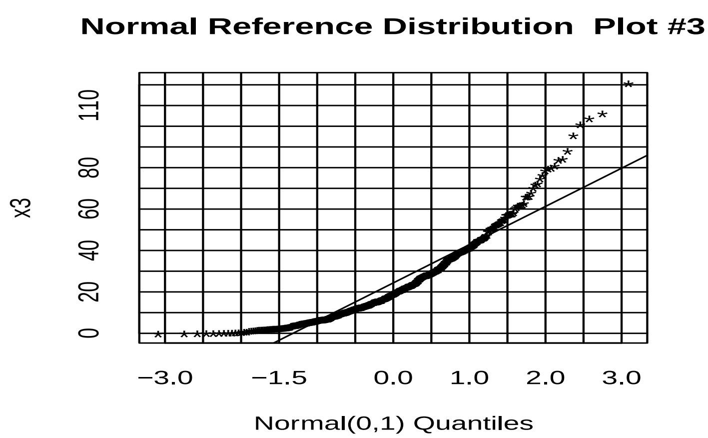
Plot 2 ___________

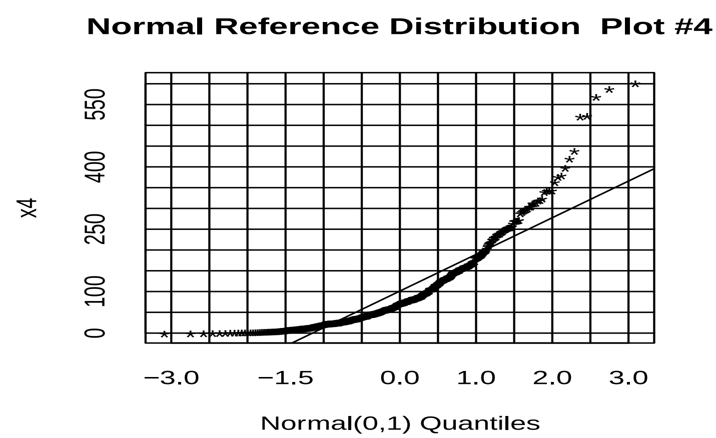
Plot 3 ___________

Plot 4 ___________









P4. (15 points) An experiment was conducted to investigate if the impact of the carcinogen DMBA could be delayed by treatment with a potential beta-blocker. Fifty mature rats of the same general health were given the beta-blocker and then were injected with DMBA. The times in days, after exposure, at which the carcinoma was diagnosed for the rats are given below. From past studies, 150 days after exposure, a carcinoma was detected in untreated rats.

10	12	13	16	37	42	43	45	55	63
66	82	99	100	100	101	107	117	122	135
138	140	142	149	150	151	154	165	170	183
194	214	218	219	224	229	232	247	268	268
298	299	325	332	379	400	434	464	499	537

Does a Weibull Distribution appear to provide an adequate fit to the data? Justify your answer using both a GOF test and a graphical plot.

P5. (25 Points) A major problem in the Gulf of Mexico is the excessive capture of game fish by shrimpers. A random sample of the catch of 50 shrimpers yield the following data concerning the catch per unit effort (CPUE) of Red Snappers, a highly sought game fish. Let C_i be the CPUE for the ith shrimper. The data, $C_1, C_2, \ldots, C_{50}$ is given next.

0.6	0.7	1.1	1.3	1.8	2.0	2.3	2.7	2.9	3.1
3.9	4.3	4.4	4.9	5.2	5.4	6.1	6.8	7.1	8.0
9.4	10.3	12.9	15.9	16.0	22.0	22.2	22.5	23.0	23.1
23.9	26.5	26.7	28.4	28.5	32.2	40.2	42.5	47.2	48.3
55.8	57.0	57.2	64.9	67.6	71.3	79.5	114.5	128.6	293.5

1. CPUE data is often modeled using a Log-Normal distribution. Does the above data appear to be from a Log-Normal distribution? Explain your answer with both a normal reference distribution plot and a GOF test.
2. Use the Box-Cox transformation of the CPUE data to determine the most appropriate power transformation to transform the CPUE distribution to Normality. How does the fit from the Box-Cox transformation compare to the fit for the log transformation?
3. Use the R program from Handout 10 (or any other program of your choice) to draw 10,000 bootstrap samples from the CPUE data. From the 10,000 samples, estimate the standard error of the sample mean for the $Y_i = log(C_i)$, data. Compare this estimate to the usual estimate $\frac{S_Y}{\sqrt{n}}$, where S_Y is the sample standard deviation computed from the $n = 50$ values of $Y_i = log(C_i)$.
4. Use your bootstrap samples to estimate the mean and standard deviation of the following sample statistics for $Y = log(C)$
   a. The sample median, $\widehat{Q}_Y(.5)$
   b. The sample standard deviation, S_Y
   c. The sample MAD, $\widehat{MAD}_Y$
5. Historically, the log(CPUE) data was modeled as a random sample from a $N(3, (1.5)^2)$ distribution. Compare your bootstrap estimates of the mean and standard deviation of $\widehat{Q}_Y(.5)$ and S_Y from part 4. of this problem to the theoretical mean and standard deviation of $\widehat{Q}_Y(.5)$ and S_Y based on $Y = log(C)$ having a $N(3, (1.5)^2)$ distribution.

P6. (25 points) A company has designed a new battery system for electric powered automobiles. To estimate the lifetime of the system, the design engineers place the batteries in 25 electric powered cars and test them under simulated city driving. Let Y_i be the time to failure of the batteries of the ith car, $i = 1, \ldots, 25$. The failure times are recorded in units of 20,000 miles. The company wants to know the probability that the sample mean based on 25 observations will estimate the true mean within a margin of error of $\pm.2$ (4000 miles), provided that the true mean has a value of 5 (100,000 miles), that is, approximate, $P[-0.2 \leq \bar{Y} - 5 \leq 0.2]$.

1. From past studies, the distribution of the time to failure of the braking system is exponential with a mean value of 5 (100,000 miles). Let $\bar{Y}$ be the sample mean time to failure of the 25 cars.
   a. What is the **exact** distribution of $\bar{Y}$ if the exponential model is still valid?
   b. What is the mean and standard deviation of $\bar{Y}$ if the exponential model is still valid?
2. Simulate 10,000 random samples of size 25 from the an exponential distribution with $\beta = 5$. Compute the sample mean from each of the 10,000 samples. Display a normal distribution reference plot for the 10,000 sample means. Does the plot suggest that the sampling distribution of $\bar{Y}$ is approximately normal?
3. Compute or estimate $P[-0.2 \leq \bar{Y} - 5 \leq 0.2]$ in each of the following manners:
   a. Using the exact distribution of $\bar{Y}$, compute $P[-0.2 \leq \bar{Y} - 5 \leq 0.2]$.
   b. Using the central limit theorem, compute $P[-0.2 \leq \bar{Y} - 5 \leq 0.2]$.
   c. Using your simulated 10,000 $\bar{Y}$s, approximate, $P[-0.2 \leq \bar{Y} - 5 \leq 0.2]$.
4. What is the level of agreement in the three computations/estimators of $P[-0.2 \leq \bar{Y} - 5 \leq 0.2]$.