STATISTICS 641 - ASSIGNMENT 6

DUE DATE: 11:59pm (CST), MONDAY, NOVEMBER 8, 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.

ASSIGNMENT 6 - DUE 11:59pm (CST), MONDAY, NOVEMBER 8, 2021

- Read Handout 11
- Supplemental Reading from Devore book: Chapters 7 & 9 & Sections 15.3
- The data for Problems I., IV. and V. are on our Google Drive in Data folder

  `Assignment06_Fall2021_data.txt`

P1. (16 Points) An experiment was designed to evaluate whether or not rainfall can be increased by treating clouds with silver iodide. Rainfall was measured from 60 clouds, of which 30 were chosen randomly to be seeded with silver iodide. The objective is to describe the effect that seeding has on rainfall. The measurements are the amounts of rainfall in acre-feet from the 60 clouds.

Seeded: 151, 450, 124, 235, 357, 110, 302, 671, 118, 115, 275, 275, 2550, 243, 201, 199, 130, 119, 92, 91, 92, 98, 1650, 1200, 1180, 900, 700, 460, 340, 330

Unseeded: 246, 268, 275, 348, 305, 311, 206, 279, 426, 269, 257, 299, 337, 329, 319, 312, 327, 342, 351, 205, 151, 426, 154, 353, 396, 441, 254, 263, 278, 281

NOTE: In what follows, you should first check whether the data are Normally distributed. If not, apply a Box-Cox transformation.

1. Place 95/95 lower tolerance intervals on the amount of rainfall amounts from both seeded and unseeded clouds. If you had to transform one or both of the datasets, create a bound for the transformed data, then back-transform to get a bound on the original scale.
2. Place 95% confidence intervals on the average rainfall from both seeded and unseeded clouds. If you had to transform one or both of the datasets, use the studentized bootstrap, because our confidence interval procedures for a mean are not appropriate for transformed data.
3. Place 95% confidence intervals on the median rainfall from both seeded and unseeded clouds. Note that, since with the Normal distribution, the median equals the mean, you can just apply confidence interval approaches for a mean. If you had to transform one or both of the datasets, go ahead and use the confidence interval approach for a mean and back-transform.
4. What can you conclude about the effect of the seeding on the amount of rainfall?

P2. (10 Points) Twenty-eight bundles of impregnated carbon fibers of length 20 mm are exposed to gradually increasing stress until they finally fail. The stress at failure are recorded as follows. The maximum stress that can be applied to the fibers is 3 and four of the fibers had **not** failed at that stress so a value of 3 was assigned to the four fibers:

2.526, 2.546, 2.628, 2.669, 2.869, 2.710, 2.731, 2.751, 2.771, 2.772, 2.782, 2.789, 2.793, 2.834,
2.844, 2.854, 2.875, 2.876, 2.895, 2.916, 2.919, 2.957, 2.977, 2.988, 3, 3, 3, 3

1. Estimate with a 95% confidence interval the average stress to failure for the carbon fibers without specifying the distribution of the stress to failure values. Do this two ways: (i) using an asymptotic CI based on the results of the R function `survfit`, and (ii) using the studentized bootstrap, treating the censored observations as true stress values (i.e., ignoring the censoring).
2. Estimate with a 95% confidence interval the average stress to failure for the carbon fibers assuming the distribution of the stress to failure values has a Weibull distribution. Do this using the parametric bootstrap. To estimate the Weibull parameters, use the `survreg` function:

```
fit <- survreg(Surv(x, delta) ~ 1, dist = "weibull")
shape_est <- 1 / fit$scale
scale_est <- exp(fit$coef)
```

P3. (9 Points) The National Institute for Standards and Technology conducted a study to develop standards for asbestos concentration. Asbestos dissolved in water was spread on a filter, and punches of 3-mm diameter were taken from the filter and mounted on a transmission electron microscope. An operator counted the number of asbestos fibers on each of 200 grid squares yielding the following counts: (the researcher no longer had the original counts just the following grouped data and the mean of the 200 counts $\bar{Y} = \frac{1}{200}\sum_{i=1}^{200} Y_i = 4940/200 = 27.7$)

	Grouped Counts							
	0-10	11-15	16-20	21-24	25-27	28-30	31 or more	Total
O_i	2	1	36	52	50	39	20	200
E_i	.12	2.43	34.62	57.51	45.36	32.62	27.34	200
$\frac{(O_i-E_i)^2}{E_i}$	30.25	.85	.06	.53	.48	1.25	1.97	35.39

1. The consulting statistician computed the Chi-square GOF and obtained $p - value < .001$ and then stated that the Poisson model provided an inadequate fit to the data. Do you agree with his results? If not correct his computations and reassess the fit of the Poisson model.
2. Assuming that the statistician was in fact correct and the Poisson model provided a reasonable fit to the counts, construct a 95% confidence interval on the average number of asbestor fibers per 3-mm diameter area. (Hint: If $Y_1, \ldots, Y_n$ are iid Poisson(λ) r.v.'s, then by the central limit theorem, the distribution of $\frac{\sqrt{n}(\bar{Y}-\lambda)}{\sqrt{\lambda}}$ is approximately $N(0,1)$ for large n.

P4. (15 Points) The space shuttle uses epoxy spherical vessels in an environment of sustained pressure. A study of the lifetimes of epoxy strands subjected to sustained stress was conducted. The data giving the lifetimes (in hours) of 100 strands tested at a prescribed level of stress is given in the following table.

.18	3.1	4.2	6.0	7.5	8.2	8.5	10.3	10.6	24.2
29.6	31.7	41.9	44.1	49.5	50.1	59.7	61.7	64.4	69.7
70.0	77.8	80.5	82.3	83.5	84.2	87.1	87.3	93.2	103.4
104.6	105.5	108.8	112.6	116.8	118.0	122.3	123.5	124.4	125.4
129.5	130.4	131.6	132.8	133.8	137.0	140.2	140.9	148.5	149.2
152.2	152.9	157.7	160.0	163.6	166.9	170.5	174.9	177.7	179.2
183.6	183.8	194.3	195.1	195.3	202.6	220.0	221.3	227.2	251.0
266.5	267.9	269.2	270.4	272.5	285.9	292.6	295.1	301.1	304.3
316.8	329.8	334.1	346.2	351.2	353.3	369.3	372.3	381.3	393.5
451.3	461.5	574.2	656.3	663.0	669.8	739.7	759.6	894.7	974.9

1. Estimate with a 99% confidence interval the probability that an epoxy strand subjected to the prescribed stress will survive for 300 hours. Use the Agresti-Coull approach.
2. Estimate with 99% certainty the time, $L_{.95}$, such that at least 95% of epoxy strands under the prescribed stress would have lifetimes greater than $L_{.95}$. You can assume that the lifetimes follow an exponential distribution.
3. Using the above data, predict with 95% certainty the lifetime of a strand subjected to the prescribed stress. Again, assume an exponential distribution.

P5. (20 Points) An experiment was conducted to determine the strength of a certain type of braided cord after weathering. The strengths of 56 pieces of cord that had been weathered for 30 days were investigated. The 56 pieces of cord were placed simultaneously in a tensile strength device. The device applies an increasing amount of force until the cord fails. The following strength readings (psi) are given below.

19.7	21.6	21.9	23.5	24.2	24.4	24.9	25.1
26.4	26.9	27.6	27.7	27.9	28.4	29.8	30.7
31.1	31.1	31.7	31.8	32.6	34.0	34.8	34.9
35.1	36.6	37.0	37.7	38.7	38.7	39.0	39.6
40.0	41.4	41.4	41.8	42.2	43.5	44.5	45.0
45.5	45.9	46.3	46.7	46.7	47.0	47.0	47.4
47.6	48.6	48.8	57.9	58.3	67.9	84.2	97.3

1. The manufacturer of the cords would like to estimate the proportion of weathered cords having tensile strength less than 50 psi. Provide a 95% confidence interval based on the information from the failure data from the 56 cords.
2. The manufacturer is planning a sales campaign to promote its cord and would like to state a tensile strength value for its cord. Provide the manufacturer with a 95% confidence interval that would provide an estimate of the median tensile strength value for the weathered braided cord.
3. In order to determine if the braided cord has tensile strength that falls within federal specifications, the manufacturer wants to determine an interval of strength values, (T_L, T_U), such that the manufacturer would be 95% confident that the interval would contain at least 90% of all strength values for its braided cords.

P6. (30 points) **Multiple Choice Questions** Select the letter of the **BEST** answer.

(1) A pipeline engineer is investigating the strength of pipe used to transport gasoline. The company has a warehouse of pipe and wants to determine an interval of values which will have 95% confidence of containing 90% of the strength readings. The engineer will construct the interval using a random sample of 275 pipes and recording their strengths: $Y_1, \ldots, Y_{275}$. A plot of the data reveals that the data is highly right skewed. The Shapiro-Wilk statistic for the transformation $X_i = \sqrt{\frac{1}{Y_i}}$ yields p-value=0.3467. Which of the following methods would be best for constructing the interval of strength values?

A. Use a studentized Bootstrap prediction interval.

B. Estimate the population pdf using a kernel density estimator and then use the MLEs of the parameters in the kernel density estimator.

C. Use a distribution-free prediction interval because $n = 275$ is very large.

D. Use the inverse of the end points of the interval constructed using the X_is: $\left(1/U_X^2, 1/L_X^2\right)$.

E. None of the above procedures would be acceptable.

(2) The bootstrap procedure for constructing a C.I. for a parameter θ is often used instead of a distribution-based procedure in constructing the C.I. when

A. the researcher wants just a rough idea of the values of θ.

B. the parametric procedure produces a very wide C.I.

C. a nonparametric and parametric procedure yield very different C.I.s.

D. when the conditions for applying the parametric procedure are not satisfied.

E. all of the above are true

(3) A study is to be conducted to estimate the mean tensile strength in newtons per square meter (N/m^2)of a new alloy. What sample size is needed to ensure that the sample mean will estimate the average tensile strength to within 10 N/m^2 with a reliability of 99%. Tensile strength generally has a normal distribution with a standard deviation of approximately 30 N/m^2.

A. 35

B. 49

C. 60

D. 538

E. cannot be determined since σ is unknown

(4) An industrial process produces piston rings having a nominal diameter of 9 cm. A 95% confidence interval for the mean diameter of the piston rings produced during July was calculated and a 95%/95% tolerance interval was calculated for the diameters of the piston rings produced during July. Which one of the following statements is true?

A. The probability is .95 that the mean diameter will fall within the confidence interval.

B. If the engineer wanted to set limits such that 95% of the output was within these limits, then the tolerance interval would be more informative than the confidence interval.

C. The width of the 95% confidence interval is generally narrower than the width of the 95%/95% tolerance interval and hence is a more precise estimator.

D. The tolerance interval for the piston diameters could be used to determine if the mean diameter for July's output was equal to 9 cm or not.

E. All of the statements A-D are false.

(5) In using the studentized bootstrap procedure to construct confidence intervals, there are two types of approximations involved in the procedure.

A. the sample size n and the level of confidence α

B. the estimation of the population cdf, F using the edf $\widehat{F}$ and the estimation of the edf using the bootstrap cdf $\widehat{F}_B$

C. the sample from the population and the sample from the resample

D. the bias and variance of the estimator

E. the estimation of the population cdf F and pdf f

(6) Suppose that $X_1, \cdots, X_n$ are highly positively correlated with a $N(\mu, \sigma^2)$ distribution. A 95% confidence interval for μ was constructed using the formula $\bar{X} \pm (t_{\alpha/2,n-1})(s/\sqrt{n})$. The true coverage probability of this confidence interval

A. is 0.95.

B. is very close to 0.95.

C. is much less than 0.95.

D. is much greater than 0.95.

E. may be greater or less than 0.95.

(7) A 95/95 tolerance interval is to be constructed for a population having pdf $f(\cdot)$. Which one of the following statements is true?

A. The tolerance interval is always wider than a 95% confidence interval for the population parameter.

B. It is necessary to specify the family for $f(\cdot)$ in order to construct the tolerance interval.

C. The normal based tolerance interval will have approximately the correct probabilities provided the sample size is large enough for the central limit theorem to be valid.

D. The distribution-free tolerance interval will generally be wider than the tolerance interval based on a specified family for $f(\cdot)$.

E. All of the above statements are true..

(8) Let $X_1,\ X_2, \ldots, X_{10}$ be iid observations from a population having pdf f. The researcher wants a 95% confidence interval on σ the population standard deviation. The researcher does not know the exact form of f but states that it is highly right skewed. What method of constructing the C.I. would you recommend?

A. use a distribution-free approach.

B. use the Box-Cox transformation.

C. apply a normal based procedures because normal based procedures are usually robust to departures from normality.

D. construct a bootstrap C.I. for σ.

E. recommend to the experimenter that more data would be required to construct the C.I.