STATISTICS 641 - ASSIGNMENT 7 - Due 5 p.m. (CST), FRIDAY - 11/9/2012

- Read Handout 11 and Sections 6.1, 6.2, 7.1.1, 7.2.1, 7.3.1, 7.4, 9.1.1, 9.3.2, 14.6 and pages 565-566 in the Textbook

- Handin the following problem :

- The data for Problems I., IV. and V. are in the file

   `Assignment07_Fall12_data.txt`

I. (16 Points) An experiment was conducted to evaluate whether or not an additive would increase the tensile strength of an alloy used in wind turbine blades. The tensile strength of 60 blades was measured, 30 blades with the additive and 30 blades without the additive. The objective is to describe the effect of the additive on the tensile strength of the alloy. The following measurements are from the 60 blades.

   With Additive: 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

   Without Additive: 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

   1. Place 95/95 lower tolerance intervals on the strength of the blades both with and without the additive.
   2. Place 95% confidence intervals on the average strength of the blades both with and without the additive.
   3. Place 95% confidence intervals on the median strength of the blades both with and without the additive.
   4. What can you conclude about the effect of the additive on the strength of the blades based on your results in parts 1.-3.?

II. (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 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.
   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.

III. (9 points) Nylon bars were tested for brittleness. Each of 280 bars was molded under similar conditions and was tested by placing a fixed stress at specified locations on the bar. Assuming that each bar has uniform composition, the number of breaks on a given bar should be Poisson distributed with an unknown a rate of breaks, λ, appearing per square inch of bar. The following table summarizes the number of breaks found on the 280 bars:

Breaks/Bar	0	1	2	3	4	5	total
Frequency	121	110	38	7	3	1	280

   Estimate using a 99% confidence interval the average number of breaks per bar for bars placed at the specified stress. (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.

IV. (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.

```
Epoxy Data:
.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
```

(A) Estimate with a 99% confidence interval the probability that an epoxy strand subjected to the prescribed stress will survive for 300 hours.

(B) 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 $L_{.95}$.

(C) Using the above data, predict with 95% certainty the lifetime of a strand subjected to the prescribed stress.

V. (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

(A) 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.

(B) 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.

(C) 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.

(D) The manufacturer of the cord is planing some major changes to their cord which will increase the tensile strength. After making the changes in the product, it will be necessary to run a large study to validate that the changes have in fact increased the tensile strength. Determine the minimum sample size necessary to be at least 95% confident that the tensile strengths obtained in the sample will cover 90% of the tensile strengths produced in the new manufacturing process.

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

(1) For sample sizes in the range 20 to 50, the question arises why should the Agresti-Coull confidence intervals be recommended in place of the Wald confidence for bounding a population proportion. Which one of the following statements is a valid justification for using Agresti-Coull?

    A. The Agresti-Coull CI has a higher coverage probability than the Wald CI.

    B. The Wald CI has a higher coverage probability than the Agresti-Coul CI.

    C. The Agresti-Coull CI has a smaller expected width compared to the Wald CI.

    D. The Agresti-Coull CI has a larger expected width compared to the Wald CI.

    E. The Agresti-Coull CI has a smaller expected width compared to the Wald CI but the Wald CI has higher coverage probability.

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

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

    E. None of the above procedures would be acceptable.

(3) 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. when the conditions for applying the parametric procedure are not satisfied.

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

    E. all of the above are true

(4) 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

(5) 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.

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

    B. 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.

    C. 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.

    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. None of the above

(6) 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 sample from the population and the sample from the resample

    C. the bias and variance of the estimator

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

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

(7) 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 much less than 0.95.

    C. is very close to 0.95.

    D. is much greater than 0.95.

    E. may be greater or less than 0.95.

(8) A 95/95 tolerance interval is to be constructed for a population having pdf $f(\cdot)$.

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

    D. 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.

    E. All of the above.

(9) Let X_1, X_2, ..., 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. recommend to the experimenter that more data would be required to construct the C.I.

    C. use the Box-Cox transformation.

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

    E. construct a bootstrap C.I. for σ.

(10) A lobbyist for the AMA wants to determine the proportion, p, of registered voters who are in favor of the Health Plan approved by the House of Representative. She is fairly certain that p is greater than 30%. The number n of registered voters than need to be surveyed such that with 95% confidence $\hat{p}$ is within .03 of p is

    A. 1068

    B. 897

    C. 267

    D. 225

    E. Impossible to determine with the given information

BONUS Question (10 points) Produce R code, similar to the code on page 31 of Handout 11, to find the values of r and s such that $(Y_{(r)}, Y_{(s)})$ is a 95% confidence interval on the 75th percentile, $Q(.75)$, based on a sample size of n=50.

CORRECTION: On page 676 in the Textbook, the author has an approximation for computing Chi-square percentiles. The correct form is

$$\chi^2_{\nu,\alpha} \approx \nu \left(1 - \frac{2}{9\nu} + z_\alpha \sqrt{\frac{2}{9\nu}} \right)^3, \text{ where } z_\alpha \text{ is the } \alpha-\text{percentile of N(0,1)}$$