STATISTICS 641 - FINAL EXAM

Student's Name ____

Student's Email Address ____

INSTRUCTIONS FOR STUDENTS:

(1) The exam consists of 9 pages including this cover page and 22 pages of Tables.

(2) You have exactly **2 Hours** to complete the exam.

(3) Record your answers to Multiple Choice Questions on Page 9.

(4) Do not discuss or provide information to anyone concerning the questions on this exam or your solutions until I post the solutions to the exam.

(5) You may use the following:

- Calculator - Your device cannot facilitate a connection to the internet or to send text messages
- Summary Sheets - **8-pages,** 8.5"x11"**, write/type/paste on both sides of the two sheets**
- *Tables for Final Exam* which are attached.

(6) Do not use any other written material except for your summary sheets and the attachments to the exam.

(7) Do not use a computer, cell phone, or any other electronic device (other than a calculator).

I attest that I spent no more than 2 Hours to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam.

Student's Signature________________________________

I. (80 points) PLACE ONE of the following letters **(A, B, C, D, or E)** corresponding to the **BEST** answer on the **Answer Sheet - Page 9** of the Exam.

(1.) A study is conducted to determine the ability of the federal government to provide security to State Department documents. A random sample of 1000 documents was selected from each of the twenty major divisions within the State Department. A security evaluation was performed on each of the 20000 documents. From these evaluations a measure of security was computed for each document. These measures were then combined into an overall assessment of the reliability of the security system. This type of study is an example of

A. a stratified random sample.

B. an observational retrospective study.

C. a double blind study.

D. a prospective case-control study.

E. none of the above

(2.) A group of concerned citizens in the USA wants to estimate the impact of a tax increase on small business owners. The group divided the USA into 4 regions and randomly selected 3 states in each region. Within each of the twelve states, 15 counties were randomly selected. In each of these counties, the Chamber of Commerce is contacted and asked to randomly select 5 small business owners in their county. The business owners are asked a series of questions which are summarized into a economic impact index. This study is an example of

A. a simple random sample.

B. a simple random cluster sample.

C. a stratified simple random sample.

D. a stratified cluster random sample.

E. a stratified multistage cluster random sample.

(3.) A metallurgist designed a study to estimate the distribution of cracks in the cooling pipes at nuclear power plants. She randomly selected 200 sections of pipes within the cooling systems at various nuclear power plants. An x-ray is taken of each pipe and the length of the largest crack is recorded. A possible probability model for C, the length of the largest crack in a randomly selected pipe, is ?

A. Binomial

B. Normal

C. Weibull

D. Hypergeometric

E. Poisson

(4.) A researcher is studying the accuracy of gasoline station pumping devices. She wants to construct a 99% confidence interval for the average error in the pumping device using a random sample of 35 stations and measuring their errors: $E_1, \ldots, E_{35}$. A plot of the data reveals that the data is highly right skewed. The Shapiro-Wilk statistic for the transformation $X_i = log(E_i)$ yields p-value=0.538. Which of the following methods would you recommend for constructing the confidence interval for the average error, μ_E?

A. Use a distribution-free confidence interval because $n = 35$ is not very large.

B. Use the inverse of the end points of the confidence interval for μ_X, that is, use $\left(e^{L_X}, e^{U_X}\right)$.

C. Use a studentized Bootstrap confidence interval.

D. Use the standard t-Based confidence interval using the $E-$data: $\bar{E} \pm t_{.005,34} S_E/\sqrt{35}$, because $n > 30$.

E. None of the above procedures would be acceptable.

(5.) A geneticist models the occurrence of defective genes in the kidney of mouse exposed to a toxin as a Poisson process with an average rate of 1 defective gene per 20 cc of kidney tissue. He wants to examine the variation in the occurrence of defective genes by simulating the number, D, of defective genes in 100 cc of kidney tissues. One such realization of D is generated by using $U = .65$ as a random observation from a Uniform on (0,1) distribution. The corresponding value of D is

A. 1

B. 5

C. 6

D. 22

E. Cannot be determined with the given information

(6.) A chemist at a large cosmetic firm designs a study to assess the toxicity of a new skin conditioner. He simultaneously feeds 100 mice a large volume of the conditioner. The time to death is recorded for the mice. The study was terminated after 30 days at which time twelve of the mice were still alive. The data from this type of study is best described as having

A. Type I censoring

B. Type II censoring

C. Right censoring

D. Left censoring

E. Random censoring

(7.) An industrial engineer has 4 potential cdf's $F_1,\ F_2,\ F_3,\ F_4$ which she is considering using in a simulation model for evaluating the production time for a newly designed conveyor system. She measured the production time for 100 runs of a prototype of the conveyor system producing $Y_1,\ Y_2,\ \ldots, Y_{100}$ which can be considered as random realizations from a continuous cdf F. Which of the following methods would be most appropriate for selecting the best estimator of F from amongst $F_1,\ F_2,\ F_3,\ F_4$?

A. Anderson-Darling (AD) GOF statistic

B. Method of Moments (MOM) Estimator

C. Maximum Likelihood Estimator (MLE)

D. Minimum Mean Square Error (MSE)

E. None of the above would be valid

(8.) A geneticist proposes the Poisson distribution as a model for the occurrence of defective genes in the kidney of mouse exposed to a toxin. The researcher records the number of defective genes in 100 cc of tissue, $Y_1, \cdots, Y_{237}$, taken from 237 mice exposed to the toxin. Which of the following statistics would be the most appropriate procedure for determining if a Poisson distribution provided a good fit to the data?

A. Anderson-Darling statistic

B. Kolmogorov-Smirnov statistic

C. Chi-square Goodness-of-Fit statistic

D. Shapiro-Wilks statistic

E. All four GOF statistics would all be appropriate in this situation

(9.) A researcher measures 29 units randomly selected from a population and then constructs a 95% confidence interval for the population mean, (12.4, 23.6). Which of the following would be the most appropriate interpretation of this interval?

A. The population mean has a 95% probability of falling between 12.4 and 23.6.

B. There is a 95% probability that the sample mean will be between 12.4 and 23.6.

C. Approximately 95% of all population means are between 12.4 and 23.6.

D. If in the future a random sample is selected then there is a 95% chance that its sample mean will be between 12.4 and 23.6.

E. None of the above are correct interpretations.

(10.) For testing $H_1 : \tilde{\mu} > 23$ based on $n = 50$ iid data values, where $\tilde{\mu}$ is the median for a symmetric but very heavy tailed distribution, the power of a level .05 Sign Test compared to the power of a level .05 t-Test would be

A. less than the power of the t-test at $\tilde{\mu} = 28$

B. greater than the power of the t-test at $\tilde{\mu} = 28$

C. nearly equal to the power of the t-test at $\tilde{\mu} = 28$ because $n = 50$ is large

D. less than the power of the t-test at $\tilde{\mu} = 23$

E. greater than the power of the t-test at $\tilde{\mu} = 23$

(11.) An educational psychologist is studying the reading ability of young children who have been in a head start program. She wants to determine if the variation in reading scores for the population of head start students is smaller than the variation in the population of children under the age of 6 who have not attended head start. It is known from the literature that the distribution of reading scores is highly right skewed for head start students and $\sigma = 12$ for non-head start students. Suppose she uses the following test: reject $H_o : \sigma \geq 12$ if $\frac{(n-1)S^2}{12^2} > \chi^2_{.95,n-1}$, where S is the standard deviation from a random sample of 500 students from the head start population of students.

A. The actual level of significance will be very close to 0.05.

B. The actual level of significance will be greater than 0.05.

C. The actual level of significance will be less than 0.05.

D. The actual level of significance will be exactly 0.05.

E. It is impossible to determine the effect of skewness on the actual level of significance.

(12.) In a study to compare the average responses of two drugs for the treatment of heart worms, the standard drug and a new proposed drug, the researchers decided to randomly assign 10 dogs to the standard drug and 40 dogs to the new drug because there was considerable existing knowledge about the effectiveness of the standard drug. From the data in the study, the sample standard deviations were computed to be $S_{standard} = 12$ and $S_{new} = 3$. If the researchers use a pooled t-Test to test the research hypothesis: $H_1 : \mu_{standard} < \mu_{new}$, what is the major consequence of using unequal sample sizes?

A. The degrees of freedom will be random.

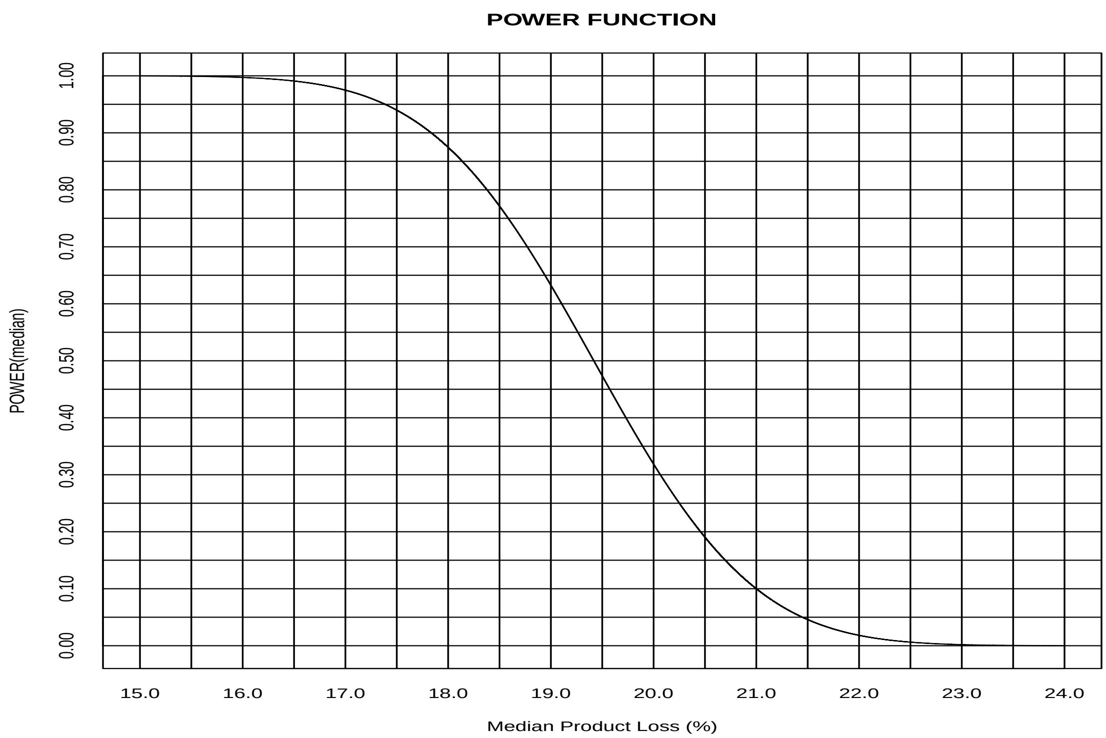
B. The significance level of the test will tend to be lower than the specified value of α.

C. The significance level of the test will tend to be higher than the specified value of α.

D. The results of the study would be invalid because the sample sizes must be equal.

E. The pooled t-Test is very robust to unequal variances so there would be very little difference from using equal sample sizes.

The following discussion will supply the information for Questions 13-14. A chemical engineer is studying the impact of a new catalyst on the product loss in a chemical production process. She knows that the median loss using the standard catalyst is 21%. She wants to determine if the median product loss is less than 21% when the new catalyst is used in place of the current catalyst. The engineer has initially planned on making 25 runs of the chemical process using the new catalyst. Prior to expending the resources on the 25 runs, she would like to determine the sensitivity of the statistical test. The following plot is the power curve for a t-Test which would address the engineer's researcher question.



Use the above description and graph to answer Questions 13-14:

(13) What is the probability of a Type II error if the median product loss using the new catalyst was 18%?

A. .875
B. .125
C. .10
D. .90
E. Cannot be determined from this graph

(14) Assuming that product loss in the chemical process has a normal distribution, the power curve for the Wilcoxon Signed Rank Test, having the same size as the t-Test, would have a plot which is

A. above the power function for the t-Test for values of $\tilde{\mu}$ less than 21 and below for values of $\tilde{\mu}$ greater than 21.
B. below the power function for the t-Test for all values of $\tilde{\mu}$.
C. above the power function for the t-Test for all values of $\tilde{\mu}$.
D. below the power function for the t-Test for values of $\tilde{\mu}$ less than 21 and above for values of $\tilde{\mu}$ greater than 21.
E. none of the above

(15.) A large corporation is evaluating two different suppliers of a raw material. The corporation's engineers obtain a random sample of 20 units from each of the suppliers and determine a crucial product characteristic from the samples: $X_1, \ldots, X_{20}$ and $Y_1, \ldots, Y_{20}$. Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be the median for each supplier's product characteristic. It is desired to test $H_o : \tilde{\mu}_1 = \tilde{\mu}_2$ versus $H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2$. An evaluation of the two data sets reveals the following:

- The width of the box in the box plot for supplier 1 is much wider than the width for supplier 2.
- The normal reference plots have the data values very close to a straight line for both data sets.
- The two data sets are independent.

The preferred test statistic is

A. Wilcoxon Rank Sum test

B. Wilcoxon signed rank test

C. Pooled t-test

D. Separate variance t-test

E. Sign test

(16.) You have been assigned to design an experiment to compare the mean responses from a placebo treatment and a new drug. You can either randomly assign n experimental units to each of the treatments or you can pair the $2n$ experimental units based on severity of the disease and then randomly assign the two treatments within each pair of experimental units. Which of the following statements is **True**?

A. Pairing results in a large increase in the power of the paired t-test provided there is negative correlation within the pairs.

B. Pairing results in a large increase in the power of the paired t-test provided there is positive correlation within the pairs.

C. Pairing would reduce the power of the t-test if the 2n experimental units have nearly the same level of severity of the disease.

D. Pairing reduces the variance of the difference in the two sample means.

E. Pairing will always increase the power of the t-test, at least to some degree.

(17.) A metallurgist for a steel company is investigation the strength of a new alloy at various times after the alloy has been heat treated. He independently heat treats 25 specimens of the alloy and records the tensile strength of the alloy every 24 hours for the 30 days following the heat treatment yielding 30 strength values for every specimen: $X_{1i}, \ldots, X_{30i}$ for each of the $i = 1, \ldots, 25$ specimens. He uses the mean responses from the 25 specimens, $\bar{X}_1, \ldots, \bar{X}_{25}$, to test the hypotheses $H_o : \mu \leq 2800$ versus $H_1 : \mu > 2800$, where μ is the average tensile strength of the alloy prior to heat treating. An evaluation of the data reveals the following:

- The Shapiro-Wilk test has p-value=0.312 for each of the 25 specimens
- For $i = 1, \ldots, 25$: $X_{1i}, \ldots, X_{30i}$ have a lag one autocorrelation values $\hat{\rho}_i$
  satisfying $\min[\hat{\rho}_1, \cdots, \hat{\rho}_{25}] > 0.89$.

A test was constructed using the following decision rule: $\frac{\bar{\bar{X}}-2800}{S/\sqrt{25}} \geq t_{.05,24}$, where $\bar{\bar{X}}$ and S are the mean and standard deviation of $\bar{X}_1, \ldots, \bar{X}_{25}$. The true level of significance of this test would be

A. very close to 0.05.

B. much less than 0.05.

C. much greater than 0.05.

D. may be greater or less than 0.05 depending on the value of σ.

E. can not be determined with the given information

(18.) An experiment is conducted to test the research hypothesis $H_1 : \theta_1 < \theta_2$ where θ_1 and θ_2 are the location parameters from two populations. The data consists of random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ with the $X_i' s$ independent of the $Y_i' s$. Box plots of the data reveals the following:

- The box plots of both the X_i's and Y_i's have more than 20% of the data values identified as outliers
- The two box plots are nearly identical in shape but the $X-$values are generally larger than the $Y-$values.

The preferred test statistic is

A. Wilcoxon Rank Sum test

B. Wilcoxon signed rank test

C. Pooled t-test

D. Separate variance t-test

E. Sign test

19.) A safety inspector of the National Highway Bureau is investigating the viability of bridges throughout the United States. In particular, she wants to determine if the design of the support beams in the bridges are related to major cracking in the support beams. Nearly all bridges have one of 6 designs used in their construction of bridge supports. For each of the 6 designs, she randomly selects 150 bridges using that design. The 900 bridges are then inspected and the number of major cracks in their support beams are recorded. What test procedure would best answer the safety inspector's question about the relationship between the six bridge support designs and the average number of major cracks per bridge?

A. Bonferroni adjusted pooled t-Test

B. Bonferroni adjusted Wilcoxon Rank Sum Test

C. Chi-square test of homogeneity of proportions

D. Hartley's F-max test

E. Breslow-Day test of homogeneity of odds ratios

(20.) An experiment is run to study the effects of DDT, a banned pesticide, on the bladder thickness of field mice. A Department of Natural Resources (DNR) researcher measured the bladder thickness of 8 mice exposed to DDT. After a 6 months exposure to DDT, the bladder thickness of the 8 mice was remeasured. From previous studies, the distribution of bladder thicknesses has been well fit by a double exponential distribution.

Mouse	1	2	3	4	5	6	7	8
PreExposure:	0.6	6.2	0.8	6.0	6.8	0.8	0.2	0.4
PostExposure:	3.3	3.4	3.6	3.5	3.4	3.5	3.4	3.6

Which test statistics would be most appropriate for determining if there is significant evidence that the average bladder thickness is smaller after exposure to DDT?

A. pooled t-Test

B. paired t-Test

C. Welch-Satterthwaite Separate Variance t-test

D. Wilcoxon Signed Rank test

E. Wilcoxon Rank Sum test

II. (20 points, 2 points each) INSTRUCTIONS Write the **ONE** number from the column on the right which **BEST** matches the statement in the column on the left. Note, there may be multiple correct responses and there may be items in the column on the right which are unused. Only **ONE** answer should be given for each statement in the column on the left.

...........A. An index of weight of evidence that data supports null hypothesis	1. Kolmogorov-Smirnov statistic
	2. Anderson-Darling statistic
...........B. A function which contains both the probability of I and II errors.	3. Chi-squared GOF statistic
	4. Shapiro-Wilk statistic
...........C. Method for evaluating differences in proportions when confounding variable is present	5. Levene-Brown-Forsythe statistic
	6. Spearman's Correlation
...........D. Technique for determining sampling distribution of C.I. pivot when n is small	7. Pearson's Product-Moment Correlation
	8. Fisher's Exact test
...........E. Method for comparing the Odds Ratios of k treatments of a disease	9. McNemar's test
	10. Cochran-Mantel-Haenszel test
...........F. An estimator of survival function when data has censored values	11. Pearson Chi-square Test
	12. Breslow-Day test
...........G. A procedure which generates k confidence intervals having a 95% overall coverage probability	13. Simpson's Paradox
	14. p-value
...........H. Test for homogeneity of $k > 2$ population variances when the population pdf's are heavily skewed	15. Probability of Type I Error
	16. Probability of Type II Error
...........I. Measure of how well a Poisson distribution fits data summarizing the occurrence of tornados	17. Power of test
	18. Specificity of test
..........J. A method of estimating parameters when pdf is specified	19. Sensitivity of test
	20. Satterthwaite Approximation
	21. Box-Cox Technique
	22. Bootstrapping
	23. Kaplan-Meier PLE
	24. Empirical distribution function
	25. Wilcoxon Rank Sum test
	26. Wilcoxon Signed Rank test
	27. Sign test
	28. Pooled t-test
	29. Welch-Satterthwaite t-test
	30. Paired t-test
	31. Confidence Interval
	32. Tolerance Interval
	33. Prediction Interval
	34. Bonferroni Interval
	35. Maximum Likelihood
	36. Method of Moments
	37. edf-based estimators
	38. Kernel Density estimator
	39. Relative Frequency Histogram
	40. Random censoring
	41. Left censoring
	42. Right censoring
	43. Type I censoring
	44. Type II censoring
	45. Non-Random censoring

ANSWER SHEET FOR MULTIPLE CHOICE QUESTIONS

Name ____

Place your answer to each of the MULTIPLE CHOICE questions in the space provided. Make sure to use UPPER case letters: A, B, C, D, E

(1.)__________ (11.)__________

(2.)__________ (12.)__________

(3.)__________ (13.)__________

(4.)__________ (14.)__________

(5.)__________ (15.)__________

(6.)__________ (16.)__________

(7.)__________ (17.)__________

(8.)__________ (18.)__________

(9.)__________ (19.)__________

(10.)__________ (20.)__________