STATISTICS 641 - EXAM II - SOLUTIONS

I. (40 points) CIRCLE ONE of the following letters **(A, B, C, D, or E)** corresponding to the best answer

(1.) **E.** Because the optimal statistic depends on the form of F_o, for example, if F_o is discrete then Chi-square GOF, if F_o is normal then S-W GOF, etc.

(2.) **B.** The cardiologist wants to have a high level of certainty that the interval contains 90% of the population hence he would require a Tolerance interval.

(3.) **C.** If $\hat{\theta}$ is a biased estimator θ with a small variance then the sampling distribution of $\hat{\theta}$ would be highly concentrated about $E[\hat{\theta}]$ which is not equal to θ, the parameter being estimated.

(4.) **D.** much less than 95% because the sampling distribution of $(n-1)S^2/\sigma^2$ would be more highly right skewed than when the population distribution has a normal distribution. This results in the true lower .025 percentile being less than the lower .025 percentile from a chi-squared distribution and the true upper .025 percentile being greater than the upper .025 percentile from a chi-squared distribution. Thus, the C.I. constructed using the percentiles from the chi-squared distribution would be too narrow to have 95% coverage.

(5.) **D.** F is a continuous cdf, use the A-D GOF statistic to select the best fitting cdf

(6.) **B :** A normal based C.I. on the mean of the transformed data and then invert back to original scale would be the best approach. Recall, the mean and median are equal for a normal distribution.

(7.) **B.** Because $Z_{.5} = 0$ on the horizontal axis corresponds to approximately 20 on the vertical axis

$Z_{.98} = 2.05$ on the horizontal axis corresponds to approximately 64 on the vertical axis.

(8.) **C.** The distribution is highly right skewed with $\hat{Q}(.5) \approx 20$ and $\hat{Q}(.98) \approx 64$.

Weibull, Exponential, and Gamma are all possible for $F(\cdot)$

Weibull(shape =1.5,scale = 25): $Q(.5) = 25 * (-log(.5))^{1/1.5} = 19.6$, $Q(.98) = 25 * (-log(1 - .98))^{1/1.5} = 62.1$

Exponential(scale=29): $Q(.5) = 29 * (-log(.5)) = 20.1$ and $Q(.98) = 29 * (-log(1 - .98)) = 113.5$

Gamma(shape =.8,scale = 25): mean is $\mu = \alpha\beta = 20 >> Q(.5)$ because distribution is right skewed.

Question II. (60 points)

(1.) (15 points) Does a LogNormal Distribution appear to provide an adequate fit to the data? Justify your answer.

- Because we are testing the normality of the data therefore use S-W test: p-value = .50 using Table A29. The closeness of the points to the line in the normal reference plot and the very large value for the p-value indicate that the normal distribution would provide an excellent fit to the distribution of log(CPUE).

(2.) (15 points) Place a 95% confidence interval on the proportion of catches which would have a CPUE less than 55, that is, on $p = P[X < 55]$

- Let B=number of $X_i < 55$. From the data, n=50, $\hat{p} = B/n = 40/50 = 0.8$.
- $n = 50 > 40$ and $n \cdot min(\hat{p}, 1 - \hat{p}) = 10 > 5$, therefore we can use the 95% Agresti-Coull C.I.:

$\tilde{B} = B + .5(1.96)^2 = 41.9208$; $\tilde{n} = n + (1.96)^2 = 53.8416 \Rightarrow$

$\tilde{p} = \tilde{B}/\tilde{n} = 41.9208/53.8416 = .7786 \Rightarrow C.I. = .7786 \pm 1.96\sqrt{(.7786)(1 - .7786)/53.8416}$

$\Rightarrow C.I. = .7786 \pm .1109 = (.6677, .88953)$

(3.) (15 points) Provide a 99% confidence interval on the median CPUE of Finfish based on the information from the 50 catches.

- $Y = log(X) \Rightarrow \; Q_Y(.5) = log(Q_X(.5)) \;\Rightarrow$

  $.99 = P[L_Y \leq Q_Y(.5) \leq U_Y] = P[L_Y \leq log(Q_X(.5)) \leq U_Y] = P[e^{L_Y} \leq Q_X(.5) \leq e^{U_Y}]$

  $Y = log(CPUE)$ has approximately a normal distribution, therefore, $\mu_Y = Q_Y(.5)$,

  A 99% C.I. on $Q_Y(.5)$ is equivalent to a 99% C.I. on μ_Y

  which is given by $\bar{Y} \pm t_{.005,49} S_Y/\sqrt{n}$

  $L_Y = 2.591 - (2.68)(1.453)/\sqrt{50} = 2.0403; \quad U_Y = 2.591 + (2.68)(1.453)/\sqrt{50} = 3.1417$

  A 99% C.I. on $Q_X(.5)$ is $\left(e^{2.0403},\; e^{3.1417}\right) = (7.69,\; 23.14)$

- A less optimal answer would be to use the interval: $(X_{(k)},\; X_{(n-k+1)})$ where k = 16 from Table VII.3

  $(X_{(16)},\; X_{(35)}) = (5.4,\; 28.5)$

(4.) (10 points) Provide the researchers with an interval of values (D_L, D_U) such that the researchers would be 90% confident that the interval would contain the CPUE for least 95% of the catches for all commercial fishermen in the Gulf of Mexico.

- $Y = log(CPUE)$ has approximately a $N(\mu_Y, \sigma_Y)$ distribution, therefore a $(P = .95, \gamma = .90)$ Tolerance interval for the distribution of Y is

  $\bar{Y} \pm K_{P,\gamma} S_Y$, where $K_{.95,\;.90} = 2.285$ for n=50 from the Tolerance Interval Table. Thus, we have

  $2.591 \pm (2.285)(1.453) = (-0.7291,\; 5.9111)$

  Therefore, a $P = .95,\; \gamma = .90$ Tolerance interval for the distribution of X is

  $(e^{-0.7291},\; e^{5.9111}) = (0.48,\; 369.11)$

  - A less optimal answer would be to use the interval:

  $(X_{(r)},\; X_{(n-s+1)})$ with $r + s = m = 1$ from Tolerance Interval Table with $P = .95$, $\gamma = .90$, and $n = 50$

  $(X_{(1)},\; X_{(51)}) = (0.6,\; \infty)$ or $(X_{(0)},\; X_{(50)}) = (0,\; 293.5)$

(5.) (5 points) The value of n such that $P[\hat{\tilde{\mu}} \leq 1.1\tilde{\mu}] = .95$ is determined as follows:

If $Y = log(X)$ then $\tilde{\mu}_Y = log(\tilde{\mu}_X)$, where $\tilde{\mu}$ is the median.

- $.95 = P[\hat{\tilde{\mu}} \leq 1.1\tilde{\mu}] = P[log(\hat{\tilde{\mu}}) \leq log(1.1) + log(\tilde{\mu})] = P[\hat{\tilde{\mu}}_Y \leq log(1.1) + \tilde{\mu}_Y]$

  Because Y is log-normal, $\tilde{\mu}_Y = \mu_Y \;\Rightarrow\; \hat{\tilde{\mu}}_Y = \hat{\mu}_Y = \bar{Y} \Rightarrow$

  $.95 = P[\bar{Y} \leq log(1.1) + \mu_Y] = P[\frac{\bar{Y}-\mu_Y}{\sigma_Y/\sqrt{n}} \leq \frac{log(1.1)}{\sigma_Y/\sqrt{n}}] = P[Z \leq \frac{log(1.1)}{\sigma_Y/\sqrt{n}}] \;\Rightarrow$

  $\frac{log(1.1)}{\sigma_Y/\sqrt{n}} = Z_{.95} = 1.645 \;\Rightarrow\; n = \frac{\sigma_Y^2(1.645)^2}{(log(1.1))^2} \approx \frac{(1.453)^2(1.645)^2}{(log(1.1))^2} = 628.9 \;\Rightarrow\; n = 629$

EXAM II SCORES:

Min = 12, Q(.25) = 65.25, Q(.5) = 71.5, Mean = 72.64, Q(.75) = 82, Max = 100