(200 pts total)

Section 1 – Short answer

(I would spend no more than 45 minutes maximum on the short answer section in order to leave yourself enough time for the long answer problems.)

1. (20 pts total; 10 pts each) (True story: Liliana Davalos needs our help! She is trying to model the data represented in the histogram below.) Below is a histogram of the number of years (recorded as an integer) between the start of a study until a patch of forest is cleared. The bins are inclusive of the number on the right of the interval (in other words, the first bin is the number of patches with Time=0 years, the second bin represents Time=1 year, the third bin represents Time=2 years, etc.)

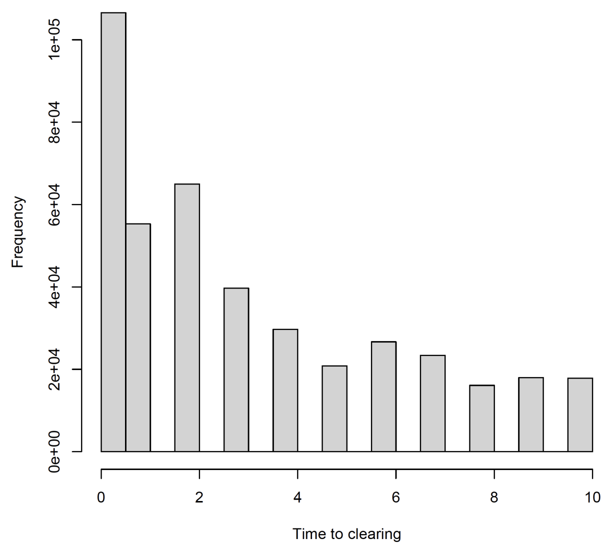


Figure 1: Number of years (integer) until a patch of forest is cleared.

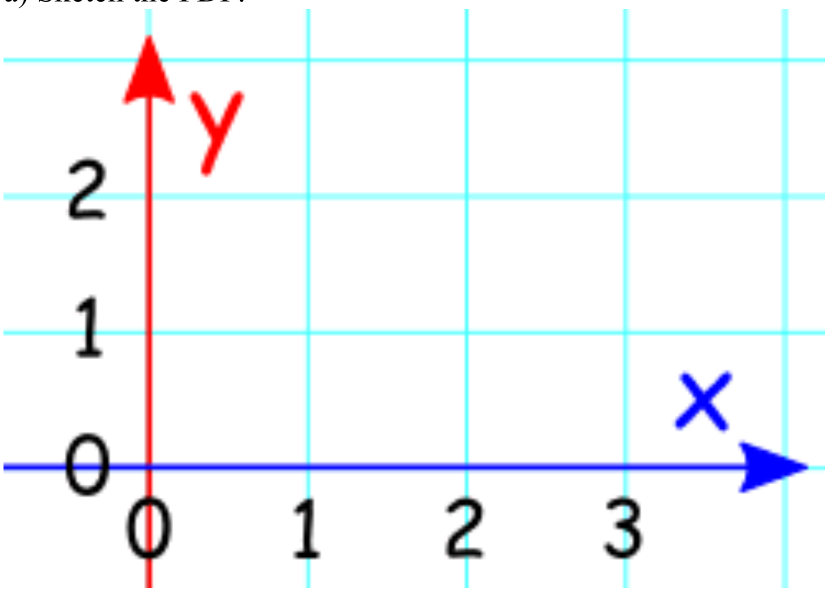
a) Name one distribution (including ballpark estimates of the distributions parameter(s)) that could have generated this dataset.

b) Name a second distribution (including estimates of the distributions parameter(s)), different from the one identified in part a, that could be used to model these data. Keep in mind that sometimes we use a distribution that know couldn't be the generating distribution but is "good enough" to use for modelling the data.

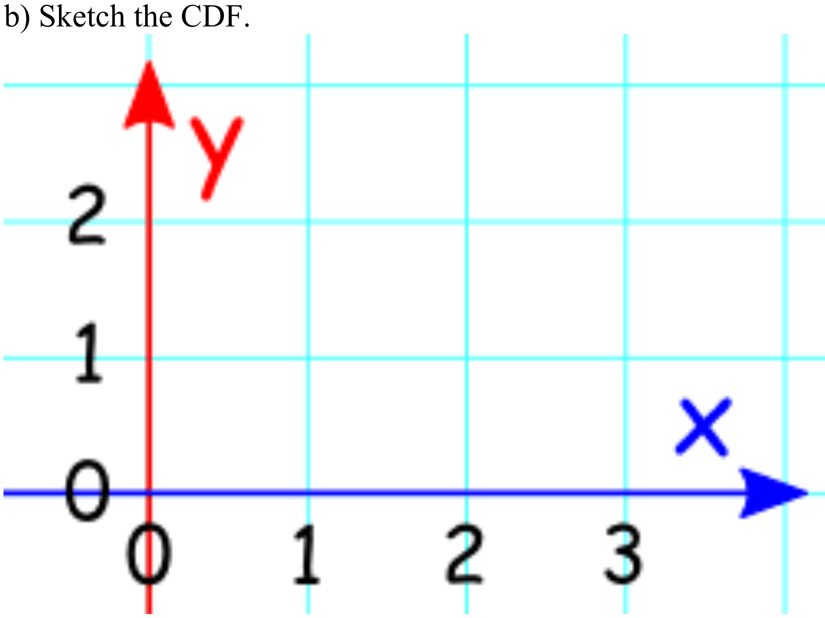
2. (20 pts total; 5 pts each)
Consider the following PDF:

$$f(x) = \begin{cases} 0.25\ for\ 0 \leq x < 2 \\ 2.0\ for\ 2 \leq x < 2.25 \\ 0\ otherwise \end{cases}$$

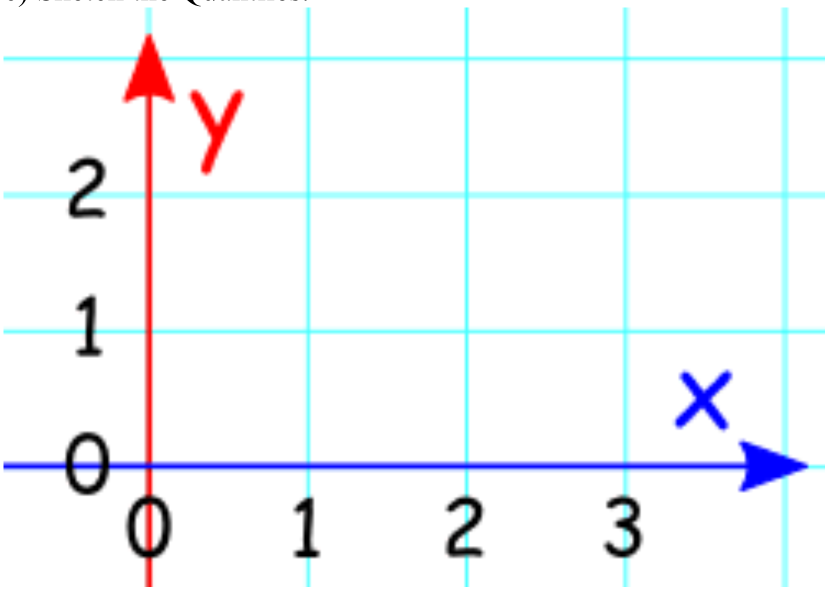
a) Sketch the PDF.



b) Sketch the CDF.



c) Sketch the Quantiles.



d) Provide a possible set of 5 random draws from this distribution.

3. (10 pts total; 2.5 pts each) Consider the following Venn diagram from Foley* et al. (2017) summarizing the number of species falling into each of the three protection regimes: 1) Being listed in Appendix I of the CITES treaty, 2) Being listed as IUCN Threatened, or 3) Being listed under the Endangered Species Act. In this question, we considered only the 1509 species depicted in this diagram. (*Foley was a former Ecology & Evolution Ph.D. student and a Biometry alumna.)

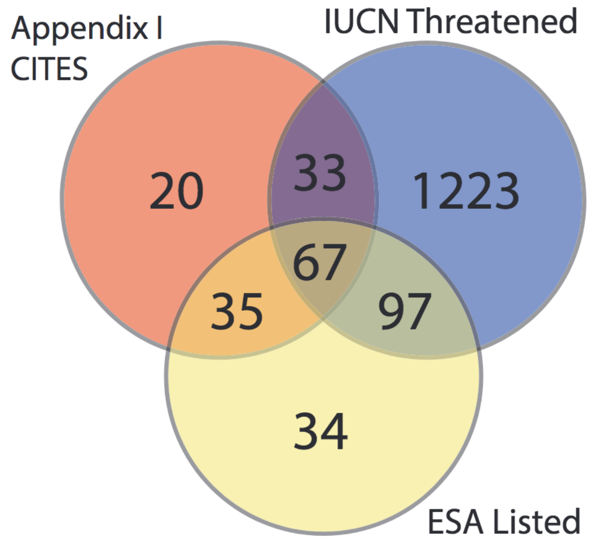


***Figure 1.** Venn diagram demonstrating the overlap between foreign bird species listed under the Endangered Species Act (ESA), foreign bird species listed under Appendix I of CITES, and foreign bird species in a threatened category as assessed by the IUCN Red List. Of the 34 listings unique to the ESA, 18 are subspecies of species not on CITES appendix I or deemed threatened by the IUCN.*

a) What is the P(CITES listed, IUCN Threatened)?

b) What is the P(ESA listed | CITES listed)?

c) What is the P(CITES listed or ESA listed)?

d) What is the marginal probability of being IUCN Threatened?

4. (10 pts) The following is a list of confidence intervals for the parameters μ, σ^2, and ${\sigma_A}^2/{\sigma_B}^2$. Circle all correct expressions. If the expression is correct only in specific circumstances, explain. Assume that the degrees of freedom represented by "[dof]" are correct.

$$P\left(\frac{ns^2}{\chi^2_{(1-\alpha/2)[n]}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{(\alpha/2)[n]}}\right) = 1-\alpha$$

$$P\left(\bar{X} + \sqrt{\frac{\sigma^2}{n}}z_{(\alpha/2)} \le \mu \le \bar{X} + \sqrt{\frac{\sigma^2}{n}}z_{(1-\alpha/2)}\right) = 1-\alpha$$

$$P\left(\bar{X} - \sqrt{\frac{s^2}{n}}t_{(1-\alpha/2)[dof]} \le \mu \le \bar{X} + \sqrt{\frac{s^2}{n}}t_{(1-\alpha/2)[dof]}\right) = 1-\alpha$$

$$P\left(\bar{X} - \sqrt{\frac{\sigma^2}{n}}z_{(\alpha/2)} \le \mu \le \bar{X} + \sqrt{\frac{\sigma^2}{n}}z_{(1-\alpha/2)}\right) = 1-\alpha$$

$$P\left(\frac{s_A^2}{s_B^2}F_{(\alpha/2)[m-1,n-1]} \le \frac{\sigma_A^2}{\sigma_B^2} \le \frac{s_A^2}{s_B^2}F_{(1-\alpha/2)[m-1,n-1]}\right) = 1-\alpha$$

$$P\left(\bar{X} + \sqrt{\frac{\sigma^2}{n}}z_{(1-\alpha/2)} \le \mu \le \bar{X} + \sqrt{\frac{\sigma^2}{n}}z_{(\alpha/2)}\right) = 1-\alpha$$

$$P\left(\bar{X} - \sqrt{\frac{s^2}{n}}t_{(\alpha/2)[dof]} \le \mu \le \bar{X} + \sqrt{\frac{s^2}{n}}t_{(1-\alpha/2)[dof]}\right) = 1-\alpha$$

$$P\left(\bar{X} - \sqrt{\frac{\sigma^2}{n}}z_{(1-\alpha/2)} \le \mu \le \bar{X} + \sqrt{\frac{\sigma^2}{n}}z_{(1-\alpha/2)}\right) = 1-\alpha$$

$$P\left(\bar{X} - \sqrt{\frac{s^2}{n}}z_{(1-\alpha/2)} \le \mu \le \bar{X} + \sqrt{\frac{s^2}{n}}z_{(1-\alpha/2)}\right) = 1-\alpha$$

$$P\left(\frac{ns^2}{\chi^2_{(\alpha/2)[n-1]}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{(1-\alpha/2)[n-1]}}\right) = 1-\alpha$$

5. (10 pts total; 5 pts each)

a) Define Type I error.

b) If a researcher accidentally used a Normal distribution as the distribution under the null hypothesis when the t-distribution was more appropriate, would the actual Type I error rate be higher, lower, or unchanged relative to the nominal (i.e. the intended) Type I error rate and why?

6. (10 pts) When conducting an analysis involving multiple hypothesis tests, we worry about inflated Type I error rates. What happens to Type II error rates when we conduct multiple comparisons? (Full credit requires deriving a mathematical expression for the "family-wise" Type II error rate across k independent hypothesis tests.)

Section 2 – Long answer

7. (40 pts) Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. random variables drawn from the Erlang distribution, whose probability density function is given by

$$f(x|k,\lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}, x \geq 0$$

a) (20 pts) Find the maximum likelihood estimator for the parameter λ.

b) (10 pts) How would you use bootstrap to find the standard error of the maximum likelihood estimator derived in part A?

c) (10 pts) A special case of the Erlang function comes about when we set $k = 1$, as the Erlang distribution becomes what is known as the Exponential distribution. With $k = 1$, find E[X].

8. (40 pts)

As part of a study into the diets of the eastern horned lizard (*Phrynosoma douglassi brevirostre*), Powell and Russell (1984,1985) investigated whether the consumption of ants varied over time. The figure below contains boxplots of dry biomass of ants collected from the stomachs of 24 adult male lizards captured in the months June-September of 1980 (24 different lizards captured in each month). *In this question, we will restrict our attention to the months of June and July only.*

a) (5 pts) What is the null hypothesis H_0 being tested? (Reminder: We are restricting our attention to June and July in this question.)

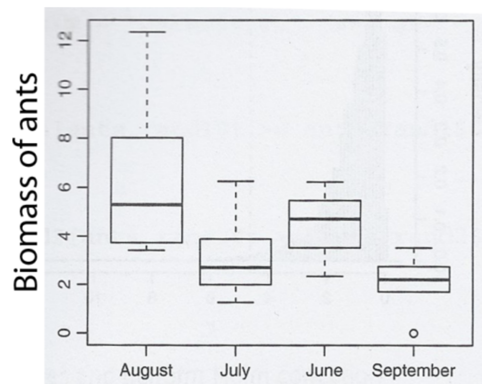


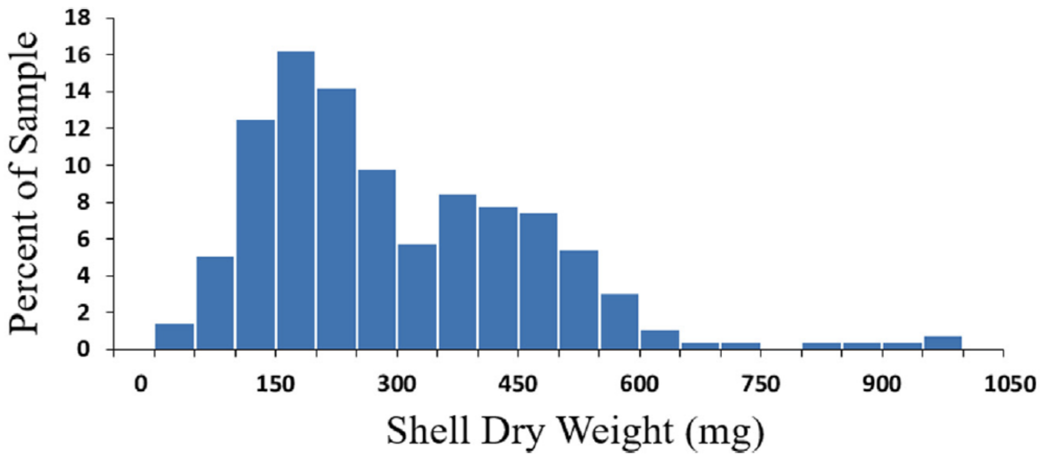
Figure 2: Boxplots of lizard consumption data.

b) (5 pts) What is the *best* parametric statistical test Powell and Russell should use to test this null hypothesis? (Here I am just looking for the full name of the test, not the equation.)

c) (15 pts) Write the equation for this statistical test. (Remember that a complete description of the statistical test includes the test statistic and its distribution under the null hypothesis.) Define all variables in your expression and calculate the degrees of freedom for any distributions used.

d) (15 pts) How might Powell and Russell have designed this study differently to increase the statistical power of their test without increasing their sample sizes? Explain the new experimental design and mathematically describe how/why this would increase the power of the test.

9. (40 points) Chase et al. (2020) studied the distribution of shell sizes among hermit crabs (*Pagurus longicarpus*) collected at West Meadow Beach. Hermit crab shell dry weight was recorded as a continuous variable, and was distributed as indicated in the histogram below.



Assume that Chase et al. (2020) modelled that data with a F distribution

$$X \sim F_{\alpha,\beta}$$

where I have labeled the two parameters of the F distribution as "α" and "β". We can use maximum likelihood to estimate the parameters of this F distribution $\hat{\alpha}$ and $\hat{\beta}$. We have two ways to express our uncertainty regarding our estimates of $\hat{\alpha}$ and $\hat{\beta}$: standard errors and confidence intervals.

a) (10 points) Describe in words the interpretation of the standard error of a parameter estimate ($\hat{\alpha}$ or $\hat{\beta}$) [Hint: The answer I'm looking for has something to do with *repeating* the experiment.]

b) (10 pts) Describe in words the interpretation of the confidence interval associated with the parameter estimates ($\hat{\alpha}$ or $\hat{\beta}$). [Hint: Once again, the answer I'm looking for has something to do with *repeating* the experiment.]

c) (20 points each = 40 pts) Briefly describe two ways to construct confidence intervals for a parameter estimate? [I'm looking for one parametric and one non-parametric approach.]