MATH 422 – SPRING 2020 – FINAL EXAM || 100 POINTS

Instructions: Show all work. No collaboration. Textbook and course notes allowed as references. Computational devices allowed.

Print Name ____

1. Consider the Weibull distribution with shape parameter $k = 2$ unknown scale parameter $\theta > 0$ which has pdf given by

$$f(x) = \frac{2x}{\theta^2} e^{-x^2/\theta^2}$$

   (a) Find the maximum likelihood estimator (MLE) for θ

   (b) Simulate Weibull data with shape $k = 2$ and scale $\theta = 5$ for a few different sample sizes.

   R code:
```
x = rweibull(n,shape=k,scale=θ)
        thetahat_mle = <insert your MLE formula>
```
   Does your MLE estimator work well? Does it become more accurate as n increases? Include three calculated MLE $\hat{\theta}$ values each for sample sizes $n = 10^2$, $n = 10^3$, $n = 10^4$, and $n = 10^5$ (so a total 12 simulated datasets).

   (c) For the Weibull distribution, the expected value is $\mathrm{E}(X) = \theta \cdot \Gamma(1 + \frac{1}{k})$. You are given that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and that $\Gamma(1+a) = a\Gamma(a)$. Find the method of moments estimator (MOME) for θ. Use this estimator again with data from the Weibull distribution with shape $k = 2$ and scale $\theta = 5$. Compare it to the MLE. Which tends to be more accurate? Include three calculated MOME $\hat{\theta}$ values each for sample sizes $n = 10^2$, $n = 10^3$, $n = 10^4$, and $n = 10^5$ (again, a total of 12 simulated datasets).

   (d) Plot a histogram of your data and overlay the Weibull pdf with the true and estimated θ values.

   R code:
```
hist(x,breaks=25,prob=TRUE,xlim=c(0,max(x)))
        curve(dweibull(x,shape=2,scale=5),from=0,to=max(x),n=500,
              col="green",xlab="x",ylab="f(x)",lwd=2,add=TRUE)
        curve(dweibull(x,shape=2,scale=thetahat_mle),from=0,to=max(x),n=500,
              col="black",xlab="x",ylab="f(x)",lwd=2,add=TRUE,lty="longdash")
        curve(dweibull(x,shape=2,scale=thetahat_mome),from=0,to=max(x),n=500,
              col="red",xlab="x",ylab="f(x)",lwd=2,add=TRUE,lty="dashed")
```
   Do the estimated pdfs fit the histogram well? Does it become more accurate as n increases? Include one plot each for simulated data with sample sizes $n = 10^2$, $n = 10^3$, $n = 10^4$, and $n = 10^5$.

   (e) Recall our large sample rule of $n \geq 30$. Generate a sample size of $n = 30$ for Weibull data with shape $k = 2$ and scale $\theta = 5$. Construct a 95% confidence interval for θ by first constructing a 95% confidence interval for $\mu = \mathrm{E}(X)$ and then transforming the interval end points appropriately. Generate 10 such datasets and their respective confidence intervals. How many of them captured the true parameter value?

2. Consider two independent normally distributed datasets $X_i \sim \mathrm{N}(\mu, \sigma^2 = 17)$ and $Y_i \sim \mathrm{N}(\mu + 10, \sigma^2 = 13)$ both of size $n = 10$. Calculate $\mathrm{P}(\overline{X} < \overline{Y})$.

3. Let $X \sim \text{Binomial}(n, \theta)$. Consider the following estimator for θ.

$$\hat{\theta} = \frac{X + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}}$$

   (a) Show that $\hat{\theta}$ is consistent.

   (b) Calculate the bias. Show that $\hat{\theta}$ is negatively biased (tendency to underestimate) when $\theta > \frac{1}{2}$, that it is unbiased when $\theta = \frac{1}{2}$, and that is is positively biased (tends to overestimate) when $\theta < \frac{1}{2}$.

   (c) Show that $P(\hat{\theta} > \frac{1}{2}) = P(X > \frac{n}{2})$.

4. Consider a quality control engineer at a drink manufacturer that wishes to estimate the variance in bottle filling for a large batch of product. The engineer tests 7 bottles and finds their filled volume in fluid ounces to be:

$$16.12, 15.97, 16.03, 15.89, 16.07, 16.02, 15.95$$

   The manufacturer desires to keep the standard deviation below 0.05 fl oz ($H_0 : \sigma^2 \leq 0.05^2$). Test the hypothesis that the product batch meets the manufacturers target at the $\alpha = 0.05$ significance level.

5. The following table contains data form a survey of Christians on their immigration status and political leaning. Use the chi-squared test for independence to see if political leaning and immigration status are independent factors.

	Immigrants	Second generation	Third generation or higher
Conservative	1,229	782	9,053
Moderate	1,453	1,118	8,606
Liberal	2,347	1,229	7,600

6. You are given a dataset including the number of confirmed COVID-19 cases per 10 thousand (CoVp10k) residents (as of Thurs 4/30) and population density (population divided by land area in square miles, PD) for all 50 US States. The data is provided below and in an excel file. You are to conduct a linear regression analysis. Include any R code or other software you use.

   (a) Construct a linear regression model on log(CoVp10k) and log(PD).

   (b) Calculate a 95% confidence interval on the slope of the regression line.

   (c) Calculate a 95% confidence interval on the correlation between log(CoVp10k) and log(PD).

   (d) Using the fact that the population density of the entire US is 94 people per square mile, predict the (log of) number of COVID-19 cases per 10k US residents with a 95% confidence interval. Does your confidence interval capture the actual number of infections per 10k residents, which is about 322 COVID-19 cases per 10k?
   (Hint: You will need to carefully go between log and linear scales here.)

	State	Population density	COVID-19 cases per 10k residents
1	Alabama	95.75	144.34
2	Alaska	1.29	48.36
3	Arizona	60.09	107.91
4	Arkansas	57.20	107.96
5	California	250.99	122.48
6	Colorado	52.61	261.46
7	Connecticut	741.15	777.42
8	Delaware	484.10	481.64
9	Florida	375.90	4674.71
10	Georgia	176.40	122.62
11	Hawaii	222.89	5.76
12	Idaho	20.00	142.64
13	Illinois	231.36	2897.78
14	Indiana	184.56	140.88
15	Iowa	55.91	105.92
16	Kansas	35.59	133.28
17	Kentucky	111.38	161.77
18	Louisiana	107.22	622.29
19	Maine	43.07	23.57
20	Maryland	614.53	1744.11
21	Massachusetts	866.64	1022.58
22	Michigan	174.68	593.11
23	Minnesota	68.96	51.13
24	Mississippi	63.79	119.55
25	Missouri	88.32	252.97
26	Montana	7.10	7.34
27	Nebraska	24.67	393.92
28	Nevada	26.32	267.70
29	New Hampshire	148.37	68.35
30	New Jersey	1207.77	8652.86
31	New Mexico	17.18	38.17
32	New York	419.28	14517.13
33	North Carolina	206.17	54.06
34	North Dakota	10.97	10.05
35	Ohio	283.61	2366.61
36	Oklahoma	56.96	30.80
37	Oregon	41.97	63.47
38	Pennsylvania	285.66	1063.99
39	Rhode Island	1010.81	67.24
40	South Carolina	162.61	577.09
41	South Dakota	11.31	47.00
42	Tennessee	160.14	1188.78
43	Texas	104.93	407.20
44	Utah	36.47	15.85
45	Vermont	67.68	26.78
46	Virginia	211.72	2690.98
47	Washington	107.75	166.09
48	West Virginia	76.59	14.57
49	Wisconsin	106.27	385.47
50	Wyoming	6.04	9.55