**Assignment 3 Template**

**Problem 1: Fill in the information below based on your data which were generated using your ID number as the seed for the random number generator.**

**n = 30 theta = 0.5282993**

**The first 10 approximate 95% confidence intervals are:**

**[,1] [,2]**

**[1,] 0.3548087 0.7118580**

**[2,] 0.4608896 0.8057770**

**[3,] 0.3210773 0.6789227**

**[4,] 0.4246923 0.7753077**

**[5,] 0.1646434 0.5020233**

**[6,] 0.3548087 0.7118580**

**[7,] 0.3210773 0.6789227**

**[8,] 0.4979767 0.8353566**

**[9,] 0.2560082 0.6106585**

**[10,] 0.3893415 0.7439918**

**Do all 10 intervals contain only values between 0 and 1? YES**

**Depending on the value of theta is it possible that some intervals will not contain only values between 0 and 1? Why or why not?**

**Since theta = 0. 5282993, it’s not quite possible that some intervals will contain only values between 0 and 1. Since we want to be 95% confident that the interval contains data, and theta is not close to either 0 or 1, it is not quite possible**

**The proportion of approximate 95% confidence intervals which contain the true value of theta = 0.9362**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.9362 is close to 0.95 but not equal because we only constructed 10 intervals which is not a large number of experiment, so it is not close to 0.95**

**The first ten 15% likelihood intervals (approximate 95% likelihood-based confidence intervals) are:**

**[,1] [,2]**

**[1,] 0.3586556 0.7025961**

**[2,] 0.4562528 0.7887964**

**[3,] 0.3276569 0.6723431**

**[4,] 0.4229261 0.7608423**

**[5,] 0.1840830 0.5096281**

**[6,] 0.3586556 0.7025961**

**[7,] 0.3276569 0.6723431**

**[8,] 0.4903719 0.8159170**

**[9,] 0.2678842 0.6095882**

**[10,] 0.3904118 0.7321158**

**Do all 10 likelihood intervals contain only values between 0 and 1? YES**

**Depending on the value of theta is it possible that some likelihood intervals will not contain only values between 0 and 1? Why or why not?**

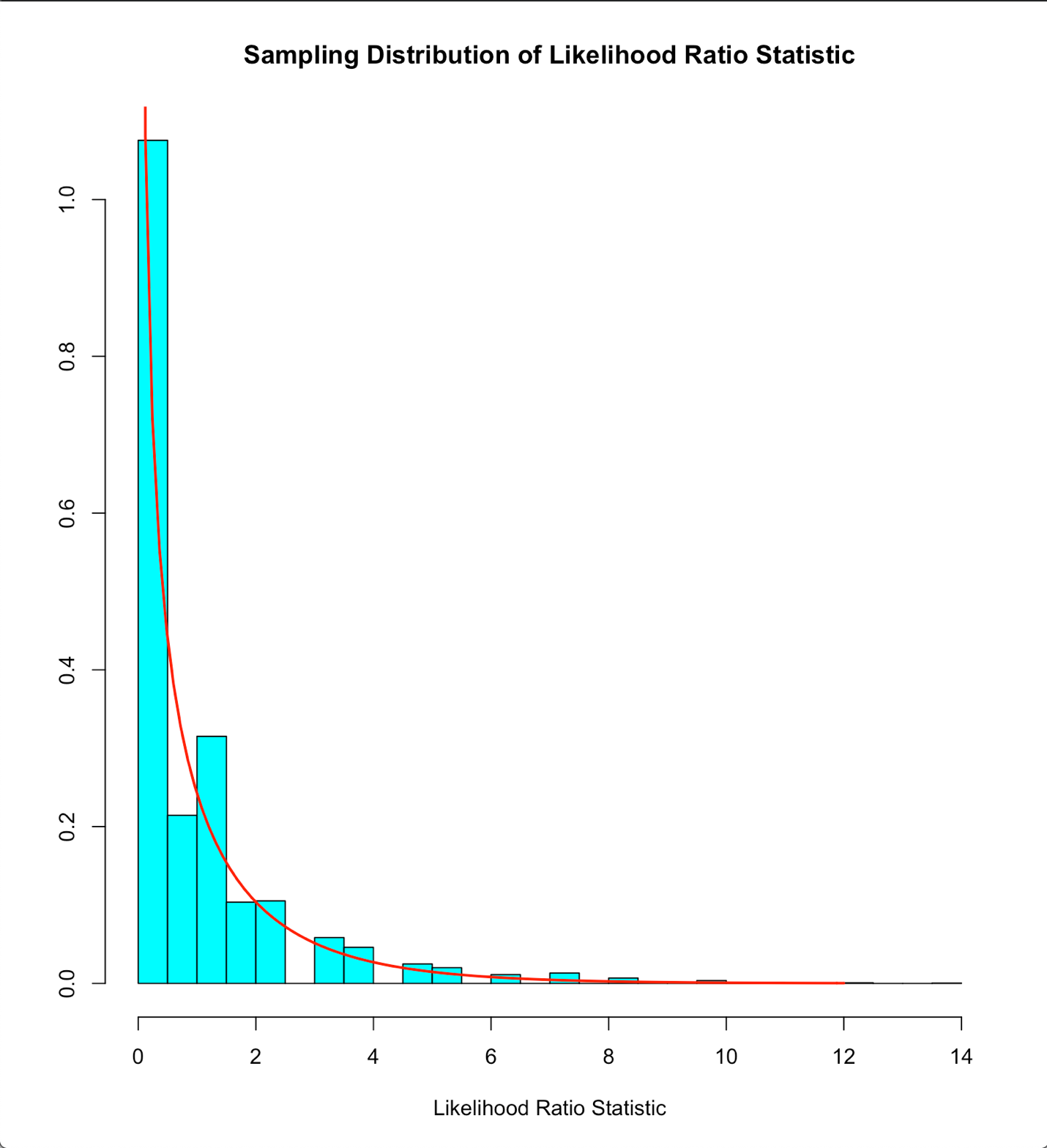
**Since theta = 0. 5282993, it’s not quite possible that some likelihood intervals will contain only values between 0 and 1. Since we want to be 95% confident that the interval contains data, and theta is not close to either 0 or 1, it is not quite possible**

**The proportion of 15% likelihood intervals which contain the true value of theta = 0.9592**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments (n) will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.9592 is close to 0.95**

**Insert the plot of the sampling distribution of the likelihood ratio statistic for n=30 here.**

****

**For Binomial data the likelihood ratio statistic is a discrete or continuous random variable?**

**Discrete random variable**

**How well does the Chi-squared(1) probability density function agree with the sampling distribution of the likelihood ratio statistic as approximated by the relative frequency histogram?**

**Chil-squared(1) probability density function agree quite well with the samepling distribution of the likelihood ratio statistic as approximated by the relative frequency histogram**

**n = 100 theta = 0.5282993**

**The first 10 approximate 95% confidence intervals are:**

**[,1] [,2]**

**[1,] 0.4120196 0.6079804**

**[2,] 0.4524912 0.6475088**

**[3,] 0.4120196 0.6079804**

**[4,] 0.3820784 0.5779216**

**[5,] 0.3920196 0.5879804**

**[6,] 0.4321766 0.6278234**

**[7,] 0.4321766 0.6278234**

**[8,] 0.3721766 0.5678234**

**[9,] 0.4627082 0.6572918**

**[10,] 0.3232625 0.5167375**

**The proportion of approximate 95% confidence intervals which contain the true value of theta = 0.9396**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.9396 is close to 0.95 but not equal because we only constructed 10 intervals which is not a large number of experiment, so it is not close to 0.95**

**The first ten 15% likelihood intervals (approximate 95% likelihood based confidence intervals) are:**

**[,1] [,2]**

**[1,] 0.4132612 0.6062142**

**[2,] 0.4527500 0.6447761**

**[3,] 0.4132612 0.6062142**

**[4,] 0.3840947 0.5769224**

**[5,] 0.3937858 0.5867388**

**[6,] 0.4329329 0.6255409**

**[7,] 0.4329329 0.6255409**

**[8,] 0.3744591 0.5670671**

**[9,] 0.4627113 0.6543121**

**[10,] 0.3267810 0.5172482**

**The proportion of 15% likelihood intervals which contain the true value of theta = 0.9396**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.9396 is close to 0.95 but not equal because we only constructed 10 intervals which is not a large number of experiments, so it is not close to 0.95**

**Insert the plot of the sampling distribution of the likelihood ratio statistic for n=100 here.**

**A picture containing screenshot

Description automatically generated**

**How well does the Chi-squared(1) probability density function agree with the sampling distribution of the likelihood ratio statistic as approximated by the relative frequency histogram?**

**Chi-squared(1) probability density function agree really well with the sampling distribution of the likelihood ratio statistic as approximated by the relative frequency histogram**

**Compare the graphs for n=30 and n=100.**

**Graph for n = 100 has data more fit with the distribution than graph for n = 30, so graph for n = 100 indicates more similarity with the chi-squared(1) distribution**

**Problem 2: Fill in the information below based on your data which were generated using your ID number as the seed for the random number generator.**

**n = 20 theta = 3**

**The first 10 approximate 95% confidence intervals are:**

**[,1] [,2]**

**[1,] 1.339708 3.430221**

**[2,] 1.797850 4.603260**

**[3,] 1.523231 3.900118**

**[4,] 1.685193 4.314811**

**[5,] 1.590644 4.072725**

**[6,] 1.980463 5.070826**

**[7,] 1.421804 3.640424**

**[8,] 1.862120 4.767818**

**[9,] 1.601686 4.100998**

**[10,] 1.987589 5.089072**

**Do all 10 intervals contain only values greater than 0? YES**

**Depending on the value of n and theta is it possible that some intervals will not contain only values greater than 0? Why or why not?**

**It is not true that depending on the value of n and theta is it possible that some intervals will not contain only values greater than 0 because intervals depend on value of theta, and theta cannot be negative, so some intervals always contain values greater than 0**

**The proportion of approximate 95% confidence intervals which contain the true value of theta = 0.9282**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.9282 is close to 0.95 but not equal because we only constructed 10 intervals which is not a large number of experiments, so it is not close to 0.95**

**The first ten 15% likelihood intervals (approximate 95% likelihood based confidence intervals) are:**

**[,1] [,2]**

**[1,] 1.588949 3.814459**

**[2,] 2.132325 5.118897**

**[3,] 1.806616 4.336992**

**[4,] 1.998710 4.798137**

**[5,] 1.886571 4.528933**

**[6,] 2.348912 5.638838**

**[7,] 1.686320 4.048208**

**[8,] 2.208552 5.301888**

**[9,] 1.899667 4.560374**

**[10,] 2.357364 5.659128**

**Do all your 10 intervals only contain values greater than 0? YES**

**Depending on the value of theta is it possible that some likelihood intervals will not contain only values greater than 0? Why or why not?**

**It is not true that depending on the value of theta is it possible that some intervals will not contain only values greater than 0 because intervals depend on value of theta, and theta cannot be negative, so some intervals always contain values greater than 0**

**The proportion of 15% likelihood intervals which contain the true value of theta = 0.951**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.951 is close to 0.95**

**Insert the plot of the sampling distribution of the likelihood ratio statistic for n=20 here.**

**A close up of a device

Description automatically generated**

**For Exponential data the likelihood ratio statistic is a discrete or continuous random variable? Continuous random variable**

**How well does the Chi-squared(1) probability density function agree with the sampling distribution of the likelihood ratio statistic as approximate by the relative frequency histogram?**

**Chi-squared(1) probability density function agree really well with the sampling distribution of the likelihood ratio statistic as approximated by the relative frequency histogram**

**Problem 3: Fill in the information below based on your data which were generated using your ID number as the seed for the random number generator.**

**mu = 3**

**sigma = 5**

**The first ten 95% confidence intervals for mu are:**

**[,1] [,2]**

**[1,] 0.70245196 4.134330**

**[2,] 0.80785124 4.267399**

**[3,] 0.86460657 4.531631**

**[4,] 1.40522116 5.012024**

**[5,] 0.74519196 4.588979**

**[6,] 1.32233670 5.542785**

**[7,] 1.19248103 4.470778**

**[8,] -0.07364781 3.848749**

**[9,] 1.00981369 4.952529**

**[10,] 1.42122798 5.152240**

**The proportion of 95% confidence intervals which contain the true value of mu = 0.9528**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.9528 is close to 0.95**

**The first ten 95% confidence intervals for sigma are:**

**[,1] [,2]**

**[1,] 3.245929 5.783065**

**[2,] 3.272100 5.829692**

**[3,] 3.468335 6.179311**

**[4,] 3.411376 6.077832**

**[5,] 3.635521 6.477175**

**[6,] 3.991773 7.111887**

**[7,] 3.100670 5.524266**

**[8,] 3.709871 6.609640**

**[9,] 3.729088 6.643879**

**[10,] 3.528856 6.287138**

**The proportion of 95% confidence intervals which contain the true value of sigma = 0.9438**

**What factors affect how close this proportion is to 0.95? Under what circumstances might you expect this proportion to be close to 0.95, and under what circumstances would you expect this proportion to be not as close to 0.95?**

**A number of experiments will affect this proportion. If a number of experiments is large, each time a 95% confidence interval for theta is constructed using data, then approximately 95% of these constructed intervals would contain the true theta. If a number of experiments is small, it will affect this proportion to be not as close to 0.95 and vice versa. In this case, theta = 0.9438 is close to 0.95**

**Problem 4: Fill in the information below based on your data which were generated using your ID number as the seed for the random number generator.**

**model = 3 mu = 3 sigma = 1.732051**

**Insert the qqplot of the data here.**

**A close up of a map

Description automatically generated**

**Based on the qqplot indicate how well the Gaussian model fits the data. Justify your conclusion.**

**Points in a qqplot given above do not lie a long a straight line which indicates that Gaussian model is not a reasonable model for these data.**

**mu0 = 4**

**Insert the output of the command** t.test(y,mu=mu0,conf.level=0.95)

**One Sample t-test**

**data: y**

**t = -5.5241, df = 29, p-value = 5.918e-06**

**alternative hypothesis: true mean is not equal to 4**

**95 percent confidence interval:**

**2.310041 3.223292**

**sample estimates:**

**mean of x**

**2.766667**

**Obtain the following information from this output:**

**value of test statistic for testing H: mu = mu0 is: -5.5241**

**degrees of freedom of t distribution = 29**

**p-value for testing H: mu=mu0 equals 5.918\*10^-6**

**95% confidence interval for mu is: [2.310041, 3.223292]**

**Insert your conclusion regarding H: mu=mu0 here.**

**Since p-value < 5%, there is a very strong evidence against Ho: mu=mu0 based on data**

**sample mean = 2.766667**

**sigma0 = 3.732051**

**sample variance = 1.495402**

**p-value for testing H: sigma = sigma0 equals 8.572076\*10^-10**

**Insert your conclusion regarding H: sigma=sigma0 here.**

**Since p-value < 5%, there is a very strong evidence against Ho: sigma = sigma0 based on data**

**95% confidence interval for sigma squared:**

**[0.9484798, 2.702466]**

**95% confidence interval for sigma:**

**[0.9738993, 1.643918]**

**Problem 5: Fill in the information below based on your data which were generated using your ID number as the seed for the random number generator.**

**alpha = 8.598549 beta = 1.320107 model = 4**

**sample correlation = 0.6343966**

**Insert the output of the command** Summary(RegModel)

**Call:**

**lm(formula = y ~ x)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-29.0532 -3.7977 0.3168 3.0681 28.0490**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 8.0222 1.7262 4.647 1.05e-05 \*\*\***

**x 1.2111 0.1491 8.124 1.38e-12 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 7.616 on 98 degrees of freedom**

**Multiple R-squared: 0.4025, Adjusted R-squared: 0.3964**

**F-statistic: 66.01 on 1 and 98 DF, p-value: 1.376e-12**

**Obtain the following information from this output:**

**estimate of the intercept = 8.0222**

**estimate of the slope = 1.2111**

**degrees of freedom of t distribution = 98**

**value of test statistic for testing H: no relationship (slope = 0) equals 8.124**

**p-value for testing the H: no relationship (slope = 0) equals 1.38\*10^-12**

**Insert your conclusion regarding the hypothesis of no relationship here.**

**Since p-value < 5%, there is a very strong evidence against the hypothesis of no relationship here (slope = 0) based on data**

**Insert the scatterplot with fitted line and plot of Residual versus x on this page.**

**A picture containing map, photo, text, sky

Description automatically generated**

**The scatterplot shows that y is proportional to x, and there is a quite strong positive relationship between x and y, or r>0**

**A screenshot of a cell phone

Description automatically generated**

**The pattern of points suggest that the points lie reasonably within a horizontal constand band around the line r\_hat\_i = 0 which suggests that the model assumptions are reasonable**

**Insert the plot of Residuals versus Muhat and qqplot of Residuals on this page.**

**A screenshot of a cell phone

Description automatically generated**

**The pattern of points suggest that the points lie reasonably within a horizontal constand band around the line r\_hat\_i = 0 which suggests that the model assumptions are reasonable**

**A close up of a map

Description automatically generated**

**Since the points in the qqplot of standardized residuals lie reasonably a long a straight line the Gaussian assumption seems reasonable. Since quantiles of the Normal distribution change more rapidly in the tails of the distribution, we expect the points at both ends of the line to lie further from the line**

**Based on the scatterplot with fitted line and the residual plots discuss the fit of the simple linear model to your data. Be sure to comment on each plot. Indicate clearly what you expect to see for each plot if the model assumptions hold and what you observe for your data.**

**Estimate of the slope: 1.21**

**95% confidence interval for the slope:**

**[4.5965240, 11.447871]**

**Estimate of mean response at x=5: 14.07772**

**90% confidence interval for the mean response at x=5:**

**[12.23896, 15.91648]**

**Estimate of predicted response at x=2: 10.44441**

**99% prediction interval for the response at x=2:**

**[-9.929617, 30.81843]**

**Estimate of sigma: 7.616284**

**95% confidence interval for the sigma:**

**[6.683017, 8.854914]**