**NOTE: Do not Turn This In. Homeworks will not be graded**

**Homework in preparation for quiz #2 and #3 – on February 19 and 26**

**Use these R exercises to develop your computer skills.**

1) Explain the following terms using the words provided in parenthesis (+0.25, each):

- Directional hypothesis vs. non-directional hypothesis (tail) =

- Binomial distribution (two parameters) =

- Poisson distribution (lambda) =

- Lognormal distribution (two parameters) =

2) Use the formula of the Binomial distribution to calculate the probability of getting a different number of successes (discrete outcomes) in ten draws, given that p(success) = 0.4 and p(failure) = 0.6. NOTE: Calculate the probability of getting zero to ten successes. There are 11 possible outcomes (I calculated the first one, the others are worth +0.10 each). Make sure the sum of the 11 probabilities adds up to 1.

Example:

P(x = 0 successes) = [ (0.6) ^ 10 ] \* number of permutations = 0.006047.

number of permutations = 10! / 10! 0! (Note 0! = 1). There is one way to get 0 successes in 10 draws.

Fill in the table below:

|  |  |  |  |
| --- | --- | --- | --- |
| Number of Successes | Probability of # Successes | Number of Permutations | Frequency of that outcome |
| 0 | 0.006047 | 1 | 0.006047 |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 |  |  |  |
| 7 |  |  |  |
| 8 |  |  |  |
| 9 |  |  |  |
| 10 |  |  |  |

SUM: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Import the file BinomialData.xlsx into R and create a histogram of this random binomial variable (showing the probability of all the different possible outcomes) (+0.250). Paste figure below, making sure you carefully consider how many bins to use:

Report the following parameters for this distribution. Use R and paste output below (+0.125 each):

* Range =
* Mean =
* Median =
* Mode =
* S.D. =

Report and interpret the following parameters for this distribution. Use R and paste output below (+0.125 each):

* Skewness =
* Kurtosis =

Question: Does this distribution look like a normal distribution? Explain why or why not ? (+0.125):

To graphically assess whether this distribution is normally distributed, create a quantile – quantile (Q – Q) plot using RCmdr. Paste the plot below and interpret what you see (+0.125):

Explain why we cannot calculate a Confidence Interval for this mean using the Z-score? (+0.250)

3) Import the file NormalData.xlsx into R and create two plots of this random normal variable (sampled from a theoretical population with a mean of 0 and a SD of 1) (+0.125 each). Create a histogram and a density plot, making sure you carefully consider how many bins to use:

Report the following parameters for this distribution. Use R and paste output below (+0.125 each):

* Range =
* Mean =
* Median =
* Mode =
* S.D. =

Report and interpret the following parameters for this distribution. Use R and paste output below (+0.125 each):

* Skewness =
* Kurtosis =

Question: Does this distribution look like a normal distribution? Explain why or why not ? (+0.125):

To graphically assess whether this distribution is normally distributed, create a quantile – quantile (Q – Q) plot using RCmdr. Paste the plot below and interpret what you see (+0.125):

Calculate the Z score for the maximum and for the minimum values of this distribution, showing how many SD units away are the sample mean they are (+0.125 each). Are you surprised of how large these standardized deviations (Z scores) are for the maximum and for the minimum observed values ?

4) Open the file (NormalSamples.xlsx) and use this information from the same random normal variable (with a mean of 0 and a S.D. of 1) for the following exercise. Analyze the values for four distributions with smaller sample sizes (n = 25, n = 50, n = 250, n = 500). Use R and paste the results below for each data sample (+0.250 for each complete row)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample Size (n) | Mean | Median | S.D. | S.E.  (show your work) | 95% CI  (show your work) |
| 25 |  |  |  |  |  |
| 50 |  |  |  |  |  |
| 250 |  |  |  |  |  |
| 500 |  |  |  |  |  |

Note: For the 95% CI, use the Z score of 1.96. Make sure you show the upper and the lower bounds.

Finally, explain your results:

* How well do the means of the four small samples (n = 25, 50, 250, 500) estimate the mean of the large population sample (n = 1000)? (+0.25)
* How does the increasing sample size affect the precision of this estimate? (+0.25)