MA311 - Introduction to Probability
Minitab Simulation Project

Name:_____ Box #:_____

1. Purpose of Project

The purposes of this project will be:

1. Gain some experience using Minitab
2. Gain some experience with simulated data
3. Gain some intuition on sums of random variables and the central limit theorem
4. Gain some intuition on transforming data

2. Work product

A typed report based on the three questions below. The report should contain

- A cover memo of submission, signed by all participants and a statement of the contributions of the team members.
- Introduction
- One section each responding to each of the three problems below.
- Any relevant graphs should be included in the report and referenced.
- Conclusion for each section.

3. Effect of sample size

Select any two distributions you like and construct the following for a random variable with those distributions. The population median is $\widetilde{\mu}$, i.e.,

$$P(-\infty < X \leq \widetilde{\mu}) = P(\widetilde{\mu} < X < \infty)$$

distribution	μ	σ	$\widetilde{\mu}$	$P(a \leq X \leq b)$

table. Now complete the two tables using randomly selected data: N = sample size, $\overline{x}$ = sample mean s = sample standard deviation, $\widetilde{x}$ = sample median. For both distributions fill in the following tables (select an appropriate a and b).

N	$\overline{x}$	s	$\widetilde{x}$	$\frac{\#\{a \leq X \leq b\}}{N}$
256				
1024				
4096				
16386				

N	$\lvert\overline{x} - \mu\rvert$	$\lvert s - \sigma\rvert$	$\lvert\widetilde{x} - \widetilde{\mu}\rvert$	$\left\lvert\frac{\#\{a \leq X \leq b\}}{N} - P(a \leq X \leq b)\right\rvert$
256				
1024				
4096				
16386				

After completing the tables draw as many conclusions as you can.

4. Transforms of variables

Let X be a uniform random variable on $[0, 1]$. Then $Y = -\ln(X)/\lambda$ is an exponential random variable. Pick suitable sample size as determined by the previous problem. Now make a histogram with 20 bins. Compute the total error between the histogram probabilities and the theoretical probabilities. If bin B_i is described by $x_i \leq Y < x_{i+1}$ then the error for this bin will be

$$\left| \frac{\{x_i \leq Y < x_{i+1}\}}{N} - P(x_i \leq Y < x_{i+1}) \right|.$$

The total error is the sum of all these probabilities. Repeat this for five selected values of λ. What do you think the effect of λ on the error is.

5. Sums of variables

Pick any distribution with mean μ and standard deviation σ. Let $X_1, \ldots, X_{36}$, be independent random variables with the given distribution. For n = 4,6,16,25,36 generate N independent samples of $X_1, \ldots, X_n$ also generate.

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

Make histograms of Z_n and draw a conclusion. For $k = -3, -2, -1, 0, 1, 2, 3$ compute the approximation of $P(Z_n \leq k)$ and compare to the standard normal distribution (errors). Make a table in n and k and draw conclusions.