Question 1 *(20 points)*

Suppose you have the opportunity to play a game with a "wheel of fortune". When you spin a large wheel, it is equally likely to stop in any position. Depending on where it stops, you win anywhere from \$0 to \$1000 (in \$1 increments, assume). Let us suppose your winnings are actually based on not one spin, but on the average of n spins of the wheel. For example, if $n = 2$, your winnings are based on the average of two spins. If the first spin results in \$580 and the second spin results in \$320, you win the average, \$450. Your objective is to use simulations to determine how does the distribution of your winnings depends on n.

To address this objective, perform a simulation in `R` using the following guidelines. Note: you need to show your work in R for all the parts, but only the summary table in part k) carries points. You won't get credit if you don't complete the summary table in part k).

a) Find the theoretical mean and theoretical standard deviation of the winnings.

b) Perform simulating 1 spin, 2 spins, 3 spins, 4 spins, 5 spins, 6 spins, 7 spins, 8 spins, 9 spins, and 10 spins. For each number of spins, perform 1,000 replications. Consider the number of spins as your sample size (n) and the number of replications as the number of samples. For example,
1,000 replications (samples): each replication (sample) containing 1 spin (sample size of 1)
1,000 replications (samples): each replication (sample) containing 2 spins (sample size of 2)
1,000 replications (samples): each replication (sample) containing 3 spins (sample size of 3)
1,000 replications (samples): each replication containing 4 spins (sample size of 4), etc.

c) Find the sample mean for each replication for each spin category. So you will have 1,000 sample means for experiment with 1 spin, 1,000 sample means for experiment with 2 spins, etc...

d) Find the mean of the sample means of each replication (sample).

e) Find the standard deviation of sample means of each replication (sample).

f) Plot a histogram for each of the 10 categories of spins. So there will be 10 histograms – one for 1 spin (1000 samples each of size 1), one for 2 spins (1000 samples each of size 2), one for 3 spins (1000 samples each of size 3), etc. Comment on how the shape of the histogram changes with increasing the number of spins (increasing sample sizes).

g) Find the theoretical standard error for each category of spins.

h) Compare the theoretical mean with the mean of sample means found in part d above.

i) Compare the theoretical standard error (part g above) with the standard deviation of sample means (part e above).

j) Find the probability of winning more than \$600 for each spin category. For example, find $P(Winning > \$600)$ with 1 spin, $P(Winning > \$600)$ with 2 spins, ... , $P(Winning > \$600)$ with 10 spins.

k) Summarize the results in the table below and comment on what you observe.

Spins (n)	1	2	3	4	5	6	7	8	9	10
Theoretical Mean										
Mean of Sample Means										
Theoretical Standard Error										
Standard Deviation of Sample Means										
$P(winning > \$600)$										

Question 2 *(10 points)*

The CSV file *SupermarketTrans* contains over 14,000 transactions made by supermarket customers over a period of approximately two years. (The data are not real, but real supermarket chains have huge data sets similar to this one.) Column A contains the date of the purchase, column B is a unique identifier for each customer, columns C–G contain information about the customer, columns H–J contain the location of the store, columns K–M contain information about the product purchased, and the last two columns indicate the number of items purchased and the amount paid.

For this question, consider this data set the population of transactions.

a) If you were interested in estimating the mean of Revenue for the population, why might it make sense to use a stratified sample, stratified by product family, to estimate this mean?

b) Suppose you want to generate a stratified random sample, stratified by product family, and have the total sample size be 250. If you use proportional sample sizes, how many transactions should you sample from each of the three product families?

c) Using the sample sizes from part b, generate a corresponding stratified random sample. What are the individual sample means from the three product families? What are the sample standard deviations?

Question 3 *(5 points)*

Major consulting firms such as Accenture, Ernst & Young Consulting, and Deloitte & Touche Consulting employ statistical analysis to assess the effectiveness of the systems they design for their customers. In this case a consulting firm has developed an electronic billing system for a trucking company in the City of Stockton. The system sends invoices electronically to each customer's computer and allows customers to easily check and correct errors. It is hoped that the new billing system will substantially reduce the amount of time it takes customers to make payments. Typical payment times—measured from the date on an invoice to the date payment is received—using the trucking company's old billing system had been 39 days or more. This exceeded the industry standard payment time of 30 days.

The new billing system does not automatically compute the payment time for each invoice because there is no continuing need for this information. The management consulting firm believes that the new system will reduce the mean bill payment time by more than 50 percent. The mean payment time using the old billing system was approximately equal to, but no less than, 39 days. Therefore, if μ denotes the new mean payment time, the consulting firm believes that μ will be less than 19.5 days. Therefore, in order to assess the system's effectiveness (whether $\mu < 19.5$ days), the consulting firm selects a random sample of 65 invoices from the 7,823 invoices processed during the first three months of the new system's operation. Whereas this is the first time that the consulting company has installed an electronic billing system in a trucking company, the firm has installed electronic billing systems in other types of companies. Analysis of results from these other companies show that, although the population mean payment time varies from company to company, the population standard deviation of payment times is the same for different companies and equals 4.2 days. The payment times for the 65 sample invoices are manually determined and are given in the CSV file named *PaymentTimes*. If this sample can be used to establish that the new billing system substantially reduces payment times, the consulting firm plans to market the system to other trucking firms.

Answer the following questions pertaining to the situation above

a) *2 points* Assuming that the standard deviation of the payment times for all payments is 4.2 days, construct a 95% confidence interval estimate to determine whether the new billing system was effective. State whether or not the billing system was effective.

b) *1 point* Using the 99% confidence interval, can we be 99% confident that the billing system was effective?

c) *2 points* If the population mean payment time is 19.5 days, what is the probability of observing a sample mean payment time of 65 invoices less than 18.1077 days?

Question 4 *(2 points)*

Toll booths on the New York State Thruway are often congested because of the large number of cars waiting to pay. A consultant working for the state concluded that if service times are measured from the time a car stops in line until it leaves, service times are exponentially distributed with a mean of 2.7 minutes. What proportion of cars can get through the toll booth in less than 3 minutes?

Question 5 *(3 points)*

Leslie loves to swim and compete in races. The time it takes Leslie to swim 100 yards in a race follows a normal distribution with mean of 62 seconds and standard deviation of 2 seconds. In her next five races, what is the probability that she will swim under a minute exactly twice?

Question 6 *(4 points)*

The Internal Revenue Service is studying the category of charitable contributions. A sample of 25 returns is selected from young couples between the ages of 20 and 35 who had an adjusted gross income of more than $100,000. Of these 25 returns, five had charitable contributions of more than $1,000. Four of these returns are selected for a comprehensive audit.

a) *2 points* What is the probability exactly one of the four audited had a charitable deduction of more than $1,000?

b) *2 points* What is the probability at least one of the audited returns had a charitable contribution of more than $1,000?

Question 7 *(4 points)*

The sales of Mercedes automobiles in the Sacramento region follow a Poisson distribution with a mean of three per day.

a) *2 points* What is the probability that no Mercedes is sold on a particular day?

b) *2 points* What is the probability that for five consecutive days at least one Mercedes is sold?

Question 8 *(6 points)*

The shoplifting sensor at a certain Best Buy Electronics store exit gives an alarm 0.5 times a minute.

a) *2 points* Find the median waiting time until the next alarm.

b) *2 points* Find the first quartile of waiting time before the next alarm.

c) *2 points* Find the $30th$ percentile of waiting time until the next alarm.

Question 9 *(10 points)*

Suppose after graduating from MSBA, you work for a survey research company. In a typical survey, you mail questionnaires to 150 companies. Some of these companies might decide not to respond. Assume that the nonresponse rate is 45%; that is, each company's probability of not responding, independently of the others, is 0.45. Suppose your company does this survey in two "waves." It mails the 150 questionnaires and waits a certain period for the responses. Assume that the nonresponse rate for this first wave is 45%. However, after this initial period, your company follows up (by telephone, say) on the nonrespondents, asking them to please respond. Suppose that the nonresponse rate on this second wave is 70%; that is, each original nonrespondent now responds with probability 0.3, independently of the others. Your company now wants to find the probability of obtaining at least 110 responses total. What is the probability (fraction of successes) of getting this required number of returns from both waves?

Question 10 *(15 points)*

It is not unusual for credit card customers to default on their credit charges. Typically customers who default share similar characteristics. Assume that the chance of defaulting on their credit charges is the same for everyone in the category. Let us look at one particular group of customers with the identical characteristics. Assume that each of these customers have a 0.07 probability of defualting on his or her current credit charges. Also assume that the total charges are normally distributed with mean of $350 and standard deviation of $100. For this question, let us say that when a customer defaults, 20% of his or her charges can be recovered. The other 80% can be written off as bad debt.

a) *5 points* What is the probability that a typical customer in this group will default and produce a write-off of more than $250 in bad debt?

b) *5 points* If there are 500 customers in this group, what are the mean and standard deviation of the number of customers who will meet the description in part a?

c) *5 points* Again assuming there are 500 customers in this group, what is the probability that at least 25 of them will meet the description in part a?

Question 11 *(50 points)*

SuperFun Toys, Inc., sells a variety of new and innovative children's toys. Management learned that the preholiday season is the best time to introduce a new toy, because many families use this time to look for new ideas for December holiday gifts. When SuperFun discovers a new toy with good market potential, it chooses an October market entry date. In order to get toys in its stores by October, SuperFun places one- time orders with its manufacturers in June or July of each year. Demand for children's toys can be highly volatile. If a new toy catches on, a sense of shortage in the marketplace often increases the demand to high levels and large profits can be realized. However, new toys can also flop, leaving SuperFun stuck with high levels of inventory that must be sold at reduced prices. So, demand is a random variable. The most important question the company faces is deciding how many units of a new toy should be purchased to meet anticipated sales demand. If too few are purchased, sales will be lost; if too many are purchased, profits will be reduced because of low prices realized in clearance sales. For the coming season, SuperFun plans to introduce a new product called Weather Teddy. This variation of a talking teddy bear is made by a company in Taiwan. When a child presses Teddy's hand, the bear begins to talk. A built- in barometer selects one of five responses that predict the weather conditions. The responses range from "It looks to be a very nice day! Have fun" to "I think it may rain today. Don't forget your umbrella." Tests with the product show that, even though it is not a perfect weather predictor, its predictions are surprisingly good. Several of SuperFun's managers claimed Teddy gave predictions of the weather that were as good as many local television weather forecasters. As with other products, SuperFun faces the decision of how many Weather Teddy units to order for the coming holiday season. Members of the management team suggested order quantities of 15,000, 18,000, 24,000, or 28,000 units. The wide range of order quantities suggested indicates considerable disagreement concerning the market potential. This is where SuperFun feels that your MSBA will add value. Having a solid background in business analytics, you are required to perform statistical analysis and the profit projections. You want to provide management with an analysis of the stock-out probabilities for various order quantities, an estimate of the profit potential, and to help make an order quantity recommendation. SuperFun expects to sell Weather Teddy for $24 based on a cost of $16 per unit. If inventory remains after the holiday season, SuperFun will sell all surplus inventories for $5 per unit. After reviewing the sales history of similar products, SuperFun's senior sales forecaster predicted an expected demand of 20,000 units with a 0.95 probability that demand would be between 10,000 units and 30,000 units.

a) *3 points* Use the sales forecaster's prediction to describe a normal probability distribution that can be used to approximate the demand distribution. State the mean and standard deviation.

b) *12 points* Compute the probability of a stock-out for the order quantities suggested by members of the management team (i.e. 15,000; 18,000; 24,000;

28,000). In other words, you need to compute the probability of stock-outs for order quantity 15,000; 18,000; 24,000; and 28,000.

c) *25 points* Compute the projected profit for the order quantities 15,000; 18,000; 20,000; 24,000; and 28,000 under three scenarios: pessimistic in which sales 10,000 units, most likely case in which sales 20,000 units, and optimistic in which sales 30,000 units. Combine your judgment and the profit/loss results to suggest the optimal order quantity. Note: It might be easier to use Excel for this part of the question. But, feel free to use `R`. I created the following table (on the next page) in Excel and used it to find the profit/loss figures. But of course, calculations can be performed in `R` as well. So please use whatever you are comfortable with. The options are Excel or R.

d) *10 points* One of SuperFun's managers felt that the profit potential was so great that the order quantity should have a 70% chance of meeting demand and only a 30% chance of any stock-outs. What quantity would be ordered under this policy, and what is the projected profit under the three sales scenarios?

See figure on the next page.

	A	B	C	D	E	F	G	H
1	Order Quantity	15,000						
2	Purchase Cost per unit	$ 16.00						
3								
4		Sales		Order Quantity	Total Cost	Total Revenue		Profit
5						@ $24.00	@ $5.00	
6		Pessimistic	10,000	15,000				
7		Likely	20,000	15,000				
8		Optimistic	30,000	15,000				
9								
10	Order Quantity	18,000						
11	Purchase Cost per unit	$ 16.00						
12								
13		Sales		Order Quantity	Total Cost	Total Revenue		Profit
14						@ $24.00	@ $5.00	
15		Pessimistic	10,000	18,000				
16		Likely	20,000	18,000				
17		Optimistic	30,000	18,000				
18								
19	Order Quantity	20,000						
20	Purchase Cost per unit	$ 16.00						
21								
22		Sales		Order Quantity	Total Cost	Total Revenue		Profit
23						@ $24.00	@ $5.00	
24		Pessimistic	10,000	20,000				
25		Likely	20,000	20,000				
26		Optimistic	30,000	20,000				
27								
28	Order Quantity	24,000						
29	Purchase Cost per unit	$ 16.00						
30								
31		Sales		Order Quantity	Total Cost	Total Revenue		Profit
32						@ $24.00	@ $5.00	
33		Pessimistic	10,000	24,000				
34		Likely	20,000	24,000				
35		Optimistic	30,000	24,000				
36								
37	Order Quantity	28,000						
38	Purchase Cost per unit	$ 16.00						
39								
40		Sales		Order Quantity	Total Cost	Total Revenue		Profit
41						@ $24.00	@ $5.00	
42		Pessimistic	10,000	28,000				
43		Likely	20,000	28,000				
44		Optimistic	30,000	28,000				

Figure 1: Profit Projections Calculations Layout on Excel for Question 11 c.