Session 1339

A Hands-On Approach to Teaching Undergraduate Engineering Students the Concept of Economic Project Risk

Edward L. McCombs and Camille F. DeYong, PhD

Industrial Engineering and Management, Oklahoma State University, Stillwater, OK 74078

Abstract

Most engineering economic analysis textbooks explain the concept of economic project risk, including methods for estimating data. However, students often do not develop an appreciation for the difficulties involved in developing estimates. The assignment discussed in this paper uses active learning to develop estimates of maintenance costs for an automobile. Students first develop estimates without any guidance. Data simulating partial historical data for maintenance costs for a rental car fleet is then created for a class exercise. The data follows a beta distribution with known parameters, although the students are unaware of this at the time. Students are provided with a histogram showing 'their' data and are asked to estimate the optimistic, pessimistic and most likely values from the graph. The mean and variance for the distribution is calculated using the common estimation equations for the beta distribution. Finally, the mean and variance of the sample data is calculated and compared to the mean and variance obtained through the estimation. This provides a clear example of the pitfalls associated with relying on an interpretation of data, or intuition, rather than using the data itself, since the estimated variance is generally radically different from the analytical variance. This exercise also provides the instructor an opportunity to discuss topics such as sampling, graphing, spreadsheet usage, optimistic/most likely/ pessimistic techniques, statistical analysis and parameter estimation.

I. Introduction

Most engineering economic analysis textbooks explain the concept of economic project risk, including methods for estimating data. However, we have found that students often do not develop an appreciation for the difficulties involved in developing estimates, including how to estimate the risk associated with the estimate. The assignment discussed in this paper accomplishes this objective by using active learning to develop estimates of maintenance costs for an automobile.

The undergraduate Engineering Economic Analysis course at Oklahoma State University is a junior level course. The only prerequisite for the course is Calculus. Students often have not yet taken probability and statistics, which makes discussions of economic risk more challenging. The instructor precedes this assignment with a lecture on engineering project financial risk and the difficulty of obtaining relevant and useful data for project analysis. This assignment is given after the assignment of the course project, which requires the students to choose between buying/leasing a car, a pickup and a sport utility vehicle. One of the costs that the students must estimate in this project is the maintenance cost for the vehicle over the planning horizon.

Students working in teams are asked to develop maintenance cost estimates without any guidance from the instructor. Results and sources for data are discussed, and students are asked how confident they are in their estimates. Since they are bright, enterprising engineering students, they are generally quite sure of their results, although they are unable to provide more than an intuitive measure.

The instructor spends some time discussing the use of pessimistic, optimistic and most likely estimates when no data is available. Basic statistical concepts and the concept of risk are introduced at this point as well.

Partial historical data representing maintenance costs for a rental car fleet is simulated through a class exercise by drawing slips of paper from a bowl. The data follows a beta distribution with known parameters, although the students are unaware of this at the time. The concept of estimating optimistic, pessimistic and most likely values is introduced, with the optimistic and pessimistic values defined as enclosing all possible values the distribution can take on. Students are provided with a histogram showing 'their' data and are asked to estimate the optimistic (least cost), pessimistic (most cost) and most likely values from the graph.

The mean and variance for the 'distribution' is calculated using the common estimation equations for the beta distribution. Since many of our students have not taken probability and statistics yet, the instructor spends time defining mean and variance and demonstrating methods of calculation.

The mean and variance of the sample data is calculated and a comparison is made between the estimated mean/variance and the sample mean/variance. This provides an example of the pitfalls associated with relying on an interpretation of data, or intuition, rather than using the data itself, since the estimated variance is generally radically different from the analytical variance. This occurs largely because of the tendency to underestimate the endpoints, H and L.

Finally, the beta distribution parameters, α and β, are calculated using the sample mean and variance. The distribution function is overlaid on top of the histogram of raw data, providing visual evidence of where the estimation process introduces error.

This exercise also provides the instructor an opportunity to discuss topics such as risk analysis, sampling, graphing, spreadsheet usage, optimistic/most likely/ pessimistic techniques, statistical analysis and parameter estimation.

II. Risk

Park and Sharpe-Bette[3, p. 356] define risk as the “situation for which outcomes are not known with certainty but about which we do have good probability information.” Risk analysis is used by decision makers to improve the decision making process. When applied properly, risk analysis can enhance the decision maker’s understanding of the risks associated with an investment alternative[4]. The most important assumption underlying risk analysis is the belief that a manager can make a better decision when he or she has an understanding of the probability distribution underlying the financial estimates.

One way to develop “good” probability information is to utilize the optimistic, most likely and pessimistic estimating procedure developed with the Project Evaluation and Review Technique (PERT)[3]. These estimates are then used to calculate the parameters for a beta probability distribution[2], which represents the probability distribution of the issue under consideration. The result of this method is that the decision maker gains knowledge of the mean, variance and the probability distribution, which are used to assist the decision maker in assessing the risk of the issue under consideration[5,6].

III. Beta Probability Distribution

The standardized Beta probability distribution (range 0 –1) is given in equation 1, where α and β are the shape parameters for the distribution[3].

$$f(x) = \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} x^{a-1}(1-x)^{\beta-1} \quad 0 \le x \le 1,\ a > 0,\ \beta > 0 \tag{1}$$

The mode, mean and variance of this distribution are given by Equations 2, 3, and 4.

$$\mathrm{m} = \mathrm{mode}_{\mathrm{Standardized}} = \frac{\alpha - 1}{\alpha + \beta - 2} \tag{2}$$

$$\mu_{\mathrm{S}} = \mathrm{mean}_{\mathrm{Standardized}} = \frac{\alpha}{\alpha + \beta} \tag{3}$$

$$\sigma_{\mathrm{S}}^2 = \mathrm{Variance}_{\mathrm{Standardized}} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{4}$$

For an arbitrary range (L, H), the following transformation is used to derive the generalized beta probability distribution (Equation 6) from the standardized distribution.

$$\mathrm{y} = \mathrm{L} + (\mathrm{H} - \mathrm{L})\mathrm{x} \tag{5}$$

where,

L = the lower limit of the range and

H = the upper limit of the range.

$$f(y)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)(H-L)^{\alpha+\beta-1}}*(y-L)^{\alpha-1}(H-y)^{\beta-1}\quad L\le y\le H \tag{6}$$

The mode, mean and variance of the generalized beta probability distribution are shown in Equations 7, 8 and 9[2], respectively.

$$\mathrm{M}=\mathrm{mode}_{\mathrm{Generalized}}=\frac{\mathrm{L}(\beta-1)+\mathrm{H}(\alpha-1)}{\alpha+\beta-2} \tag{7}$$

$$\mu_{\mathrm{G}}=\mathrm{mean}_{\mathrm{Generalized}}=\mathrm{L}+(\mathrm{H}-\mathrm{L})\frac{\alpha}{\alpha+\beta} \tag{8}$$

$$\sigma_{\mathrm{G}}^{2}=\mathrm{Variance}_{\mathrm{Generalized}}=(\mathrm{H}-\mathrm{L})^{2}\frac{\alpha\beta}{(\alpha+\beta)^{2}(\alpha+\beta+1)} \tag{9}$$

The parameters α and β are estimated using the process described in the next section.

IV. Estimating the parameters of the Beta probability distribution

We will discuss two methods that are used to estimate the shape parameters (α, β) of a beta probability distribution. The first method relies on experience and intuition. The second method relies on incomplete historical data.

IV.1. Using experience and intuition

Historical records, experience and/or consensus are used to estimate the most likely outcome, the most optimistic outcome and the least likely outcome of the issue under consideration[1]. For example, if the issue is future maintenance costs, then the most optimistic outcome corresponds to the smallest cost that can possibly occur. The most pessimistic outcome corresponds to the largest cost that can possibly occur. The most likely outcome corresponds to the maintenance cost that will be likely to occur more often than any other value. Therefore, the most optimistic outcome represents the lower bound, L, of the proposed beta probability distribution. The most pessimistic outcome represents the upper bound, H, of the proposed beta probability distribution and the most likely outcome represents the mode, M, of the proposed beta probability distribution.

The mean of the proposed distribution is estimated using equation 10.

$$\mathrm{Mean}_{\mathrm{est}}=\frac{\mathrm{L}+4\mathrm{M}_{\mathrm{est}}+\mathrm{H}}{6} \tag{10}$$

Park and Sharpe-Bette[2] report that McBridge and McClelland show that the greatest percent difference between this approximation for the mean and the exact value of the associated beta probability distribution is 18.8%.

To simplify the computations, the estimated mean is standardized using Equation 11.

$$\text{Mean}_{\text{Std}} = \frac{\text{Mean}_{\text{est}} - \text{L}}{\text{H} - \text{L}} = \frac{\alpha}{\alpha + \beta} \tag{11}$$

The upper bound (H) and the lower bound (L) correspond to the distribution's standardized upper bound of 1 and the distribution's standardized lower bound of 0, respectively. The parameters, α and β, are calculated by assuming the standard deviation of the standardized beta probability distribution is equal to 1/6 and solving the following two simultaneous equations[1].

$$\text{Mean}_{\text{Std}} = \frac{\alpha}{\alpha + \beta} \tag{12}$$

$$\sigma^2_{\text{Std}} = \frac{1}{6^2} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{13}$$

Many textbooks omit solving for the parameters of the beta probability distribution[3,4] but we maintain that, if data can be assumed to follow a beta distribution, which is often an assumption when using optimistic, pessimistic and most likely techniques, an estimate of the shape parameters provides more complete information.

IV.2. Using incomplete data

If a manager (and other company resources) has sufficient (~100 data points) [1] data, and desires to fit the data with a beta probability distribution, then the manager can use the following method to estimate the mean, variance and probability distribution. First, the manager visually estimates the mode and calculates the mean and variance of the known data using the common equations for the sample mean and variance. The manager now uses the assumption that the standard deviation for the standardized probability distribution is equal to 1/6 and equations 9 and 10 (repeated here as equations 14 and 15, respectively), to calculate the endpoints of the generalized beta probability distribution (two equations and two unknowns).

$$\text{Variance}_{\text{sample}} = \frac{(\text{H}_{\text{calc}} - \text{L}_{\text{calc}})^2}{36} \tag{14}$$

$$\text{Mean}_{\text{sample}} = \frac{\text{L}_{\text{calc}} + 4\text{M}_{\text{est}} + \text{H}_{\text{calc}}}{6} \tag{15}$$

Now that estimates of the endpoints of the beta probability distribution are known, the manager can use the same procedure as in the previous section to estimate the shape parameters α and β.

The next section discusses an actual classroom example of using the beta probability distribution to help students understand engineering economic risk. This assignment is given to the students after the Instructor discusses the Risk Analysis topic of an Engineering Economic Analysis course.

V. Classroom Example

This assignment is designed to show undergraduate Engineering Economics students how decision makers in industry should, and should not, develop estimates of costs when they have incomplete information. The purpose of this assignment is to foster student participation in a class discussion on risk analysis. A careful reading of the following assignment will show that the mean and variance forecasts is a confounding of above two methods (using the histogram to generate the forecasts estimates for the optimistic, pessimistic and most likely values) and will likely produce estimates for the variance that is significantly less than that of the true value. While we do not generally try to mislead students, we thing that in this age of too much information, the situation described in this assignment can occur later in actual practice unless the students are not forewarned about the potential problems.

V.1. The Setup

An Excel spreadsheet is used to generate a histogram, consisting of 100 data points, from a beta probability distribution. In this assignment, we use a Beta probability distribution with parameters $\alpha = 2$ and $\beta = 6$ **(**B(2,6)**),** mean = \$10,000 and the range = \$10,000. Figure 1 below shows the histogram.

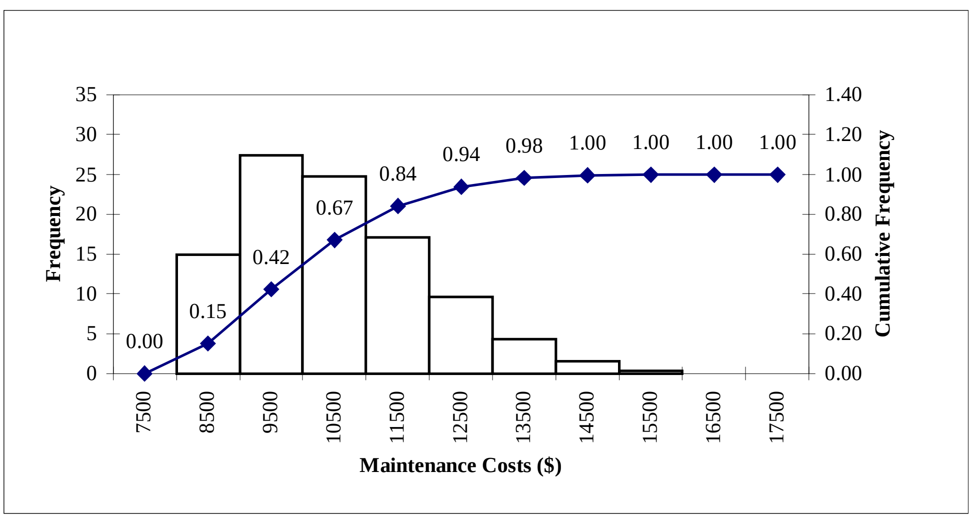

Figure 1. 100 Point Histogram for Classroom Example

The cost associated with each of the 100 data points is printed on a 1” x 1” slip of paper. Each slip of paper is then folded and placed in a container.

The students are informed that the slips of paper in the container show last year's individual maintenance cost for each of the existing 100 automobiles in a rental fleet. The students are then asked to individually draw, without replacement, one slip of paper from container and record the maintenance cost, shown on the slip, on a data collection sheet (See Appendix A). In this example, there were 55 students in the class. Therefore, 55 slips of paper were drawn. After all the students have sampled the container (without replacement), the instructor collects the data collection sheet and has an assistant use the 55 samples to generate a histogram and a cumulative frequency chart. The students are now told that, while they had the potential to gather maintenance costs for all 100 machines, they only had the resources to collect the maintenance costs for 55 autos. The effect is that the outliers for the beta distribution have a very small (or non-existent) probability of being selected in the sample, leading students to underestimate the true value for H and overestimate the true value for L.

V.2. The Assignment

Each student is given a copy of the data collection sheet, histogram, and cumulative frequency chart. The handout states that the students should assume that the maintenance costs follow a beta probability distribution. The instructor spends part of the class period discussing basic statistics, including mean, variance and probability distributions and then illustrates the use of H and L to estimate the mean and variance of the beta distribution. It is critical to carefully define H and L as the highest and lowest values that could ever occur and M as the value that will occur most frequently.

Step 1 - Students estimate the most optimistic maintenance cost (L), the most likely maintenance cost (M) and the most pessimistic maintenance cost (H) based on the data given on the data collection sheet, histogram and the cumulative frequency chart.

Step 2 – Students use the estimation equations to calculate the estimated mean and variance.

Step 3 – Students calculate the sample mean and variance of the maintenance costs shown on the data collection sheet.

Step 4 – Students compare and contrast the mean and variance estimates obtained in step 2 with those obtained in step 3. MAJOR POINT – MEAN (EST) AND MEAN(SAMPLE) WILL BE CLOSE. VARIANCE (EST) AND VARIANCE (SAMPLE) WILL NOT BE CLOSE. THIS PROVIDES THE INSTRUCTOR WITH AN OPPORTUNITY TO EXPLORE WHY THE VARIANCE IS SO FAR OFF (FOR A HEAVY TAILED DISTRIBUTION, IT IS DIFFICULT TO ESTIMATE H AND L FROM THE HISTOGRAM, CAUSING THE VARIANCE TO BE UNDERESTIMATED). POINT?? RELY ON THE SAMPLE MEAN AND VARIANCE WHEN YOU HAVE DATA. DISCUSS THREE IMPORTANT ELEMENTS OF RISK ANALYSIS: MEAN, VARIANCE AND THE PROBABILITY DISTRIBUTION. WE HAVE THE MEAN AND THE VARIANCE, BUT NEED THE DISTRIBUTION.

Step 5 – The instructor/students analytically solve for the beta distribution boundaries (H and L) using Equations 14 and 15.

Step 6 – The instructor/students use the H and L values from step 5 to iteratively solve for the beta distribution shape parameters, α and β. Use α and β and the mode estimate from step 1 (M_{est}) to develop the distribution. The distribution function is then overlaid on top of the histogram of raw data.

VI. Results

This section discusses the results of this assignment when given to a class of 55 undergraduate engineering students in a junior level Engineering Economics course. Ninety-two percent of the students chose $8,000 as the lower bound (most optimistic), 86% of the students chose $14,000 as the upper bound (most pessimistic) and 78% of the students chose a mode between $9,000 and $10,500 (most likely). The true lower and upper bounds and the mode were $7,500, $17,500 and $9,167, respectively. The estimated, sample mean and variance values are compared to the theoretical values in Table 1.

Table 1. A comparison of the student estimated mean and variance values to the theoretical mean and variance values.

		% Difference
Theoretical Mean	10,000	--
Theoretical Variance	2,083,333	--
Sample Mean	10,292	3%
Sample Variance	2,358,249	13%
Average Estimated Mean	10,447	4%
Average Estimated Variance	978,350	-53%
Range of Estimated Means	9,667 – 11,667	(-3%) – (-17%)
Range of Estimated Variances	444,444 – 1,562,500	(-79%) – (-25%)

VI.1 Discussion of Results

Before the assignment given, each student was asked to complete a self evaluation regarding their current ability to estimate cost data when they have no historical data and when they have historical data. Over 95% of the students responded that they were 'confident' or 'very confident' of their estimate of the mean obtained without historical data. Over 85% of the students responded that they were 'confident' or 'very confident' about their ability to estimate the mean using historical data. However, when questioned about specific methods, the most common approaches were to use the internet (no historical data) and to find the average (with historical data).

After the assignment, students were asked to re-assess their abilities to estimate data with and without historical data. The results were virtually the same numerically, but the students included comments about needing a measure of variation and/or risk in their responses. These comments were non-existent in the first survey. The students discovered that they need to estimate both the mean and the variance of cash flows to understand the financial risk of an investment. In addition, they acquired a methodology to make quantitative estimates when historical data is available and when it is not.

Finally, the students discovered that, without qualified expert knowledge of the system or complete data, it is difficult to accurately estimate the endpoints of a beta probability distribution. More importantly, they discovered how inaccurate endpoint estimation affects the estimated variance, which in turn affects the amount of risk, perceived. The lesson learned by the students is that it is better to rely on the data alone, unless they have specific system knowledge.

VII. Conclusion

The methodology presented in this paper helps undergraduate Engineering students understand economic project risk and highlights the potential problems a manager can face when trying to develop reasonable estimates of engineering costs. This paper illustrates two methods to forecast the theoretical mean and variance for a beta distribution. Table 3 below presents a summary of these equations.

Table 3. Sumary sheet of Equations.

Estimated	Mean	$\text{Mean}_{\text{estimated}} = \frac{L_{\text{estimated}} + 4\text{Mode}_{\text{estimated}} + H_{\text{estimated}}}{6}$
	Variance	$\text{Variance}_{\text{estimated}} = \frac{(H_{\text{estimated}} - L_{\text{estimated}})^2}{36}$
	Mode	Estimated from histogram
Sample	Mean	$\text{Mean}_{\text{sample}} = \frac{\sum_{i=1}^{n} x_i}{n}$
	Variance	$\text{Variance} = \frac{\sum_{i=1}^{n} (x_i - \text{Mean}_{\text{sample}})}{n-1}$
	Mode	Estimated from histogram
Theoretical	Mean	$\text{Mean}_{\text{Theoretical}} = L_{\text{known}} + (H_{\text{known}} - L_{\text{known}}) \frac{\alpha_{\text{known}}}{\alpha_{\text{known}} + \beta_{\text{known}}}$
	Variance	$\text{Variance}_{\text{Theoretical}} = \frac{(H_{\text{known}} - L_{\text{known}})\alpha_{\text{known}}\beta_{\text{known}}}{(\alpha_{\text{known}} + \beta_{\text{known}})^2(\alpha_{\text{known}} + \beta_{\text{known}} + 1)}$
	Mode	$\text{Mean}_{\text{Theoretical}} = L_{\text{known}} + (H_{\text{known}} - L_{\text{known}}) \frac{\alpha_{\text{known}} - 1}{\alpha_{\text{known}} + \beta_{\text{known}} - 2}$

The students overwhelmingly felt that this was an interesting and useful assignment. Their most common comments were that they now had a better understanding of the connection between variation and risk and that they were surprised that there was a quantitative method available to estimate the mean and the variance even if they had no historical data. We believe that this assignment was worthwhile and will continue to use it in future undergraduate classes.

Bibliography

1. Case, Kenneth E., 2001, Regents Professor Industrial Engineering and Management, Oklahoma State University, Discussions on minimal data set size necessary to estimate an underlying distribution, 8 January.
2.Greer, Wilis J., Jr., 1970, "Capital Budgeting Analysis with the Timing of Events Uncertain," The Accounting Review, January, 103-114.
3. Park, Chan S., and Sharpe-Bette, Gunter P., 1990, Advanced Engineering Economics, John Wiley and Sons, New York.
4. White, John A., Case, Kenneth E., Pratt, David B., and Agee, Marvin H., 1998, Principles of Engineering Economic Analysis, John Wiley and Sons, New York.
5. Young, Donovan, and Contreras, Lois E., 1975, "Expected Present Worths of Cash Flows Under Uncertain Timing," The Engineering Economist, 20(4), 257-268.
6. Zinn, C.D., and Lesso, W. G., 1977, "A Probabilistic Approach to Risk Analysis in Capital Investment Projects," The Engineering Economist, 22(4), 239-260.

Biographical Information

CAMILLE FRYE DEYONG

Dr. Camille Frye DeYong is an Assistant Professor of Industrial Engineering and Management at Oklahoma State University. She is a graduate of OSU (B.S., M.IE. and Ph.D). She has authored six technical papers and one book chapter. Dr. DeYong is an ASQ Certified Quality Engineer and served on the Eugene Grant Award Committee for ASEE. She is a member of IIE, ASQ, and ASEE.

E. L. MCCOMBS

Edward McCombs is currently a doctoral student at Oklahoma State University. He has a BS in Mechanical Engineering and a MS in Industrial Engineering. His research interests are Engineering Economics, Quality and Predictive Maintenance. Before returning to graduate school he was an Officer and Instructor for the U.S. Naval Nuclear Power School and the Engineering Manager for a secondary aluminum processor.

Appendix A

Vehicle Number	Maintenance Costs		Vehicle Number	Maintenance Costs
1	10000		36	10000
2	11000		37	11000
3	12000		38	11000
4	9000		39	11000
5	10000		40	9000
6	9000		41	9000
7	9000		42	11000
8	9000		43	9000
9	11000		44	11000
10	13000		45	10000
11	12000		46	9000
12	8000		47	11000
13	9000		48	8000
14	11000		49	11000
15	9000		50	9000
16	10000		51	9000
17	13000		52	11000
18	13000		53	10000
19	12000		54	14000
20	8000		55	11000
21	11000		56	
22	10000		57	
23	10000		58	
24	14000		59	
25	10000		60	
26	8000		61	
27	12000		62	
28	10000		63	
29	10000		64	
30	10000		65	
31	9000		66	
32	9000		67	
33	9000		68	
34	13000		69	
35	11000		70	

Appendix B

Scenario: You are an intern for a corporation that has a fleet of company cars. Your boss has decided to buy herself a new company car and needs to estimate the maintenance cost prior to purchase. Your boss knows that you took an economic analysis class at the university, so she has asked you to help with the analysis. She wants you to estimate the average maintenance cost and provide her with some measure of how sure you are about the average, so she asked you to also give her something called the variation.

This new auto is the same brand and model as the existing 100 autos currently in service, so you decide to visit your friends in the fleet maintenance department and just ask them for these figure and you'll be done! However, much to your chagrin, you discover that, for various reasons, almost half of the data has not been recorded. After wailing and moaning, you accept the data they do have, which represents the yearly maintenance cost on 55 of the 100 cars.

Your friend helps you draw a bar graph of the data (he calls it a histogram) and you spend some time looking at the data. Then you decide to use a system that has worked well for you when you didn't have any data. Your company uses this system when for making time estimates for projects. It's called the Project Evaluation Review Technique (PERT). You obtain the company PERT manual and read about making estimates. The manual says *"In the development of a PERT analysis, the decision maker provides only three subjective estimates (optimistic, most likely and pessimistic). These estimates represent the best reasonable outcome, the most likely outcome and the worst reasonable outcome that can be expected."* You think, *Wahoo!* and go back to the Histogram of your data (Figure B.1) to help you estimate the most maintenance cost that can reasonably be expected (pessimistic), the least maintenance cost that can reasonably be expected (optimistic), and the most likely maintenance cost that can reasonably be expected. The only thing that bothers you is that you have only part of the data……….

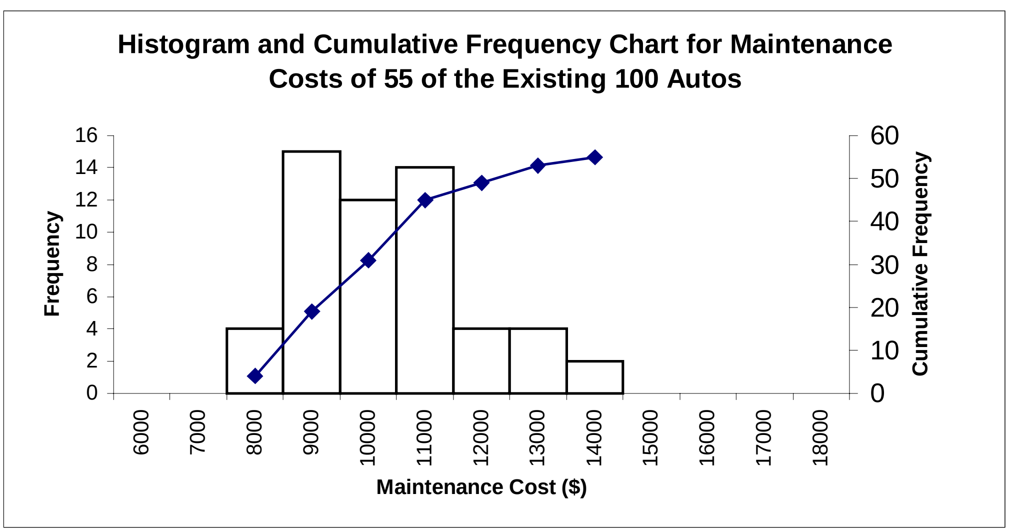

Figure B.1. Histogram and Cumulative Frequency Chart for 55 Autos

Step 1 – Estimate the optimistic, most likely and pessimistic values.
You spend a little time thinking about this, but then you decide the optimistic value will be the least maintenance cost expected (L), and the pessimistic value will be the most maintenance cost expected (H). You already knew that the most likely value is M.

H_{est} = ____________

L_{est} = ____________

M_{est} = ____________

Step 2 – Calculate the mean and variance using your estimates from Step 1 and the following formulas.

$$Mean_{est} = (L+4M+H)/6$$

$$Variance_{est} = (H-L)^2/6^2$$

$Mean_{est}$ ____________

$Variance_{est}$ ____________

Step 3 – Your friend reminds you that you could also compute the mean and variance of the data itself and offers to help you do this. You tell her – 'I already did that,' but she insists on using the following formulas:

$$\text{Variance}_{\text{sample}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = 2{,}358{,}249 \quad \text{dollars}^2$$

$$\text{Mean}_{\text{sample}} = \frac{\sum_{i=1}^{n} x_i}{n} = 10{,}292 \quad \text{dollars}$$

$Mean_{sample}$ = ____________

$Variance_{sample}$ = ____________

Step 4 - Compare your estimates for mean and variance. Are they the same? If not, why are they different? What will you tell your boss??????

	Estimated	Sample
Mean	___________________	___________________
Variance	___________________	___________________

NOTE!! At this point the instructor or the students should analytically solve for α and β using Appendix C and show students the beta probability distribution overlaid on top of the histogram. The following points should then be made.

1. The upper boundary of the distribution should be equal to L. Ask students what their estimate of L would be now that they see the distribution. Most will admit they underestimated L.

2. If no historical data is available, optimistic, pessimistic and most likely estimates obtained from an expert provides reasonable estimates.

3. If historical data is available, calculate the sample mean and sample variance of the data. If finite boundaries can be assumed and the true distribution is unknown, assume the data follows a beta distribution and use the procedure in Appendix C to obtain the shape parameters of the distribution.

Appendix C

Sometimes it is useful to estimate the maintenance costs using a probability distribution. Assume that the maintenance costs follow a beta probability distribution. The parameters for this probability distribution are the shape parameters α and β. Estimate the parameters using the method shown below.

a. Estimate the upper and lower bounds (H and L) of the beta probability distribution by estimating the Mode (Most likely value) of the maintenance cost data (the histogram might be useful for this). Use this estimated Mode, the sample mean and variance values given in #3 and the following two equations to solve for H and L.

$$\text{Variance}_{\text{sample}} = \frac{(\text{H}_{\text{calc}} - \text{L}_{\text{calc}})^2}{6^2}$$

$$\text{Mean}_{\text{sample}} = \frac{\text{L}_{\text{calc}} + 4\text{Mode}_{\text{estimated}} + \text{H}_{\text{calc}}}{6}$$

b. Calculate the standardized mean using the calculated distribution bounds in 4a.

$$\text{Mean}_{\text{std}} = \frac{\text{Mean}_{\text{sample}} - \text{L}_{\text{calc}}}{\text{H}_{\text{calc}} - \text{L}_{\text{calc}}}$$

c. Solve the following simultaneous equations for the beta probability distribution shape parameters α and β. Note that the assumption is made that the standardized variance is equal to $(1/6)^2$.

$$\text{Mean}_{\text{std}} = \frac{\alpha}{\alpha + \beta}$$

$$\text{Variance}_{\text{Std}} = \frac{1}{6^2} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

One way to solve these equations is to use trial and error, and Excel.

	A	B
1	Alpha =	Input guess here
2	Mean_{Std} =	Input standardized mean value here
3	Beta =	=(b1 – b1*b2)/b2
4	$\text{Variance}_{\text{Std}}$ =	1/36
5	Check value =	=((b1*b3)/((b1+b3)^2*(b1+b3+1)))-b4

BE CAREFUL when inputting the above formulas into Excel.

Continue to guess alpha until the check value is within 0.001 of zero. You can assume that alpha is >0 and that alpha does not have to be an integer. Record your alpha, beta, and the visually estimated Mode on the data collection sheet. In addition, record your trial values for alpha and the associated beta and check values

IEM 3503/3513
Engineering Economic Analysis

Data Collection Sheet

Mode =
α =
β =

$$\text{Variance}_{\text{Calculated from data}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = 2{,}358{,}249 \quad \text{dollars}^2$$

$$\text{mean}_{\text{Calculated from data}} = \frac{\sum_{i=1}^{n} x_i}{n} = 10{,}292 \quad \text{dollars}$$

Trial #	alpha	beta	check value
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			