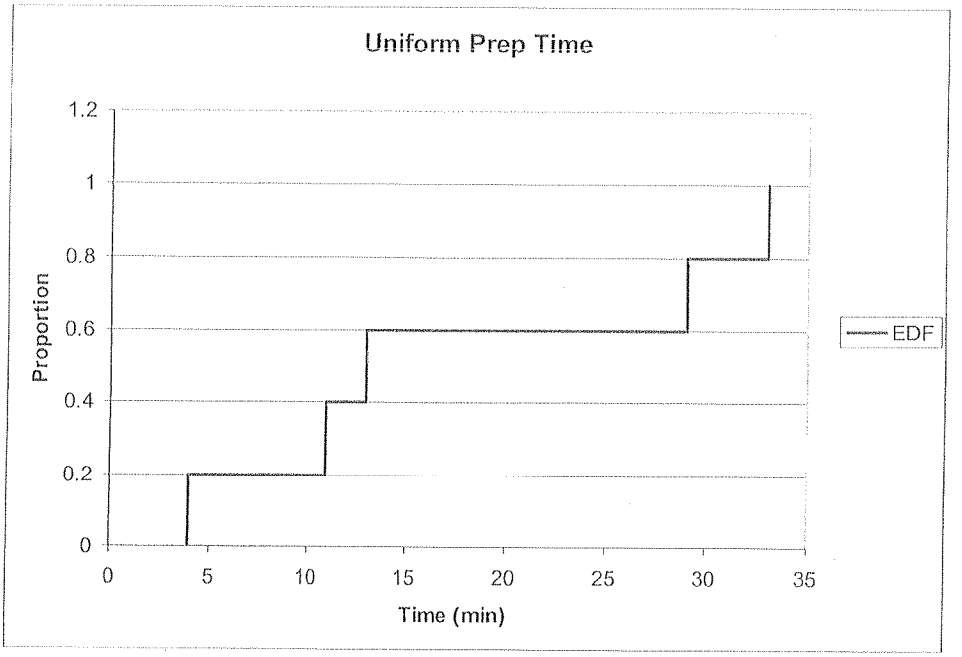
SOLUTION

MA 206 Suggested Problems
Lesson 2: Empirical Distribution Functions (EDF)

1. Suppose we obtain the following sample data on the amount of time (in minutes) that a cadet spends preparing his uniform for inspection each day:

S={11, 33, 4, 13, 29}

a. Sketch (by hand) the EDF for this data set on the following graph:



b. Using appropriate interval notation, completely specify the step-function graphed above.

$$F_5(x) = \begin{cases} 0 & x < 4 \\ 0.2 & 4 \le x < 11 \\ 0.4 & 11 \le x < 13 \\ 0.6 & 13 \le x < 29 \\ 0.8 & 29 \le x < 33 \\ 1 & x \ge 33 \end{cases}$$

c. Based on the EDF above, what proportion of cadets spend less than 10 minutes preparing their uniforms each day? Less than 25 minutes? More than 30 minutes?

Less than 10 minutes: 0.20 or 20%

SOLUTION

Less than 25 minutes: 0.60 or 60%

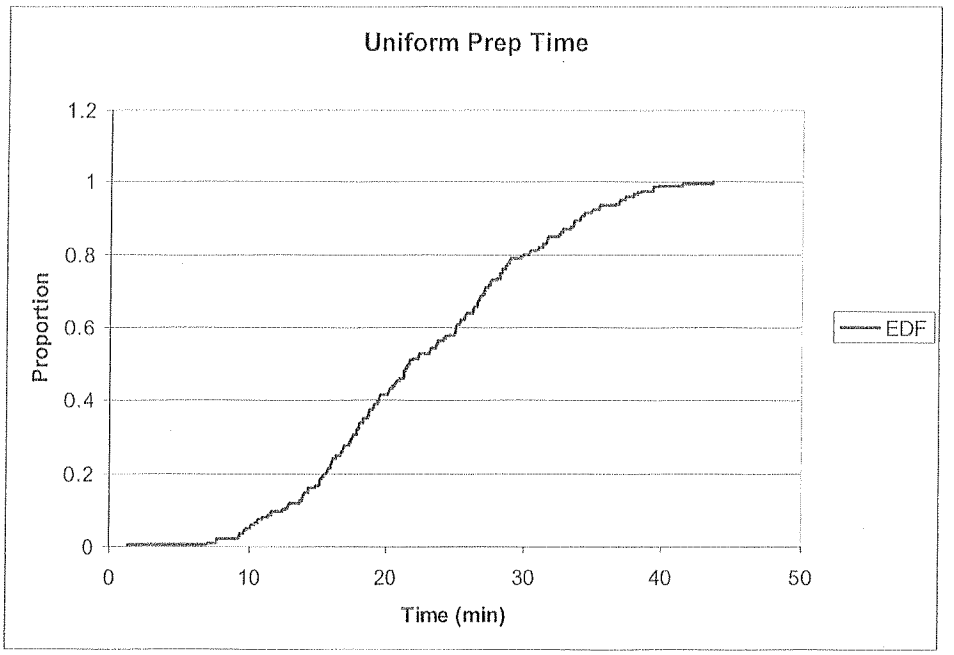
More than 30 minutes: 1-0.80 = 0.20 or 20% (since 80% of the observations are less than 30, 20% of the observations must be greater than 30)

d. Considering the original sample, do your results make sense? Why or why not?

Less than 10 minutes: 20% makes sense since 1 out of the 5 sample values is less than 10 (the third observation, which is equal to 4).
Less than 25 minutes: 60% makes sense since 3 out of the 5 sample values are less than 25 (4, 11, and 13).

More than 30 minutes: 20% makes sense since 1 out of the 5 sample values is greater than 30 (the second observation, which is equal to 33).

2. Given the same scenario as above, but using the much larger sample data set provided on the course webpage ("*Data4*" -- 200 observations), create another EDF.



a. Comment on why the new EDF looks vastly different than the EDF in Problem 1a.

We created the EDF above using 200 sample observations, whereas we created the EDF in Problem 1a using only 5 observations. Therefore, the height of each step in the new EDF is equal to 1/200 (versus 1/5 for the EDF in problem 1a). Both EDFs are step functions, but the steps in the new EDF are so small (due to the larger sample size) that they are not as readily apparent.

SOLUTION

b. Using the new EDF, what proportion of cadets spend less than 10 minutes preparing their uniforms each day? Less than 25 minutes? More than 30 minutes?

Less than 10 minutes: 0.05 or 5%

Less than 25 minutes: 0.59 or 59%

More than 30 minutes: 1-0.80 = 0.20 or 20% (since 80% of the observations are less than 30, 20% of the observations must be greater than 30)

c. Are you more, less, or equally confident in your answers above (compared to the answers you determined in Problem 1b)? Explain.

More confident! With a much larger sample size (more information), we expect the EDF to more accurately depict the amount of time cadets spend preparing their uniforms for inspection.

d. It seems that as the sample size increases, the EDF looks more like a smooth continuous function and less like a step-function. How could you use this fact to more efficiently answer the questions posed in Problems 1b and 2b?

We could model the EDF with a continuous function. If we could find a function that closely fits the EDF, we could answer the questions posed in Problems 1b and 2b by direct substitution rather than analyzing a graph or searching through a table of values.