

Lesson 27. Input Data Analysis, cont.

1 Last time

Example 1. The Midville Manufacturing Company has 2 planers for performing two different types of jobs, A and B. Each planer can process 1 job at a time. The time required to perform each job depends largely upon the number of passes that must be made: according to the planer operators, most jobs require 5 or fewer passes on a planer, and a pass takes about 10 minutes. Unfortunately, the planer department has had a difficult time keeping up with its workload. You have been asked to investigate the effect of obtaining 1 additional planer.

You have collected data for 1 week of the planer department's operation. The shop operates 8 hours per day. You have created a simulation model that accurately represents the flow of material through the shop, but you need the distributions for the interarrival times for each job type, and the service times for each job type on each planer.

- The histogram of service times matches with the planer operators' observations
- Stat::Fit wasn't able to fit the service times of type A jobs on planer 1 to any known distributions

2 Distributions directly from data

- Let's create distribution based on the frequency of service times of type A jobs on planer 1
- Based on the planer operators' observations, let's use the FREQUENCY function to put the service times into intervals $(x - 5, x + 5]$ for $x = 10, 20, 30, 40, 50, 60, 70, 80$
- We'll use the % observations in interval $(x - 5, x + 5]$ as the probability that the service time is x
- We can do something similar for the other 3 combinations of jobs and planers
- Do the resulting probability distributions for all combinations of jobs/planers look similar?

3 Pooling distributions and the two-sample Kolmogorov-Smirnov test

- Can we treat the planer service times as identical?
 - Is the service time distribution for type A jobs the same on planer 1 and planer 2?
 - Is the service time distribution for type B jobs the same on planer 1 and planer 2?
- If so, it might be reasonable to treat a third planer the same as planer 1 and planer 2
- The **two-sample Kolmogorov-Smirnov (KS) test** may be used to test for large differences in two samples
- The cdf of the empirical distribution for a sample size n is defined as

- Let $F_n(x)$ and $G_m(x)$ be the cdfs of the empirical distributions of the two samples
- The null hypothesis H_0 is that the two samples come from the same distribution
- The test statistic is defined as

- The p -value is

- Let's compute the two-sample K-S statistic in Excel between the service times of type A jobs on planer 1 and planer 2
- What can we conclude about the service time observations of type A jobs on planer 1 and planer 2?