Construction Operation Simulation

Lecture #8

Simulation model input analysis

Amin Alvanchi, PhD

Construction Engineering and Management

LinkedIn Instagram WebPage

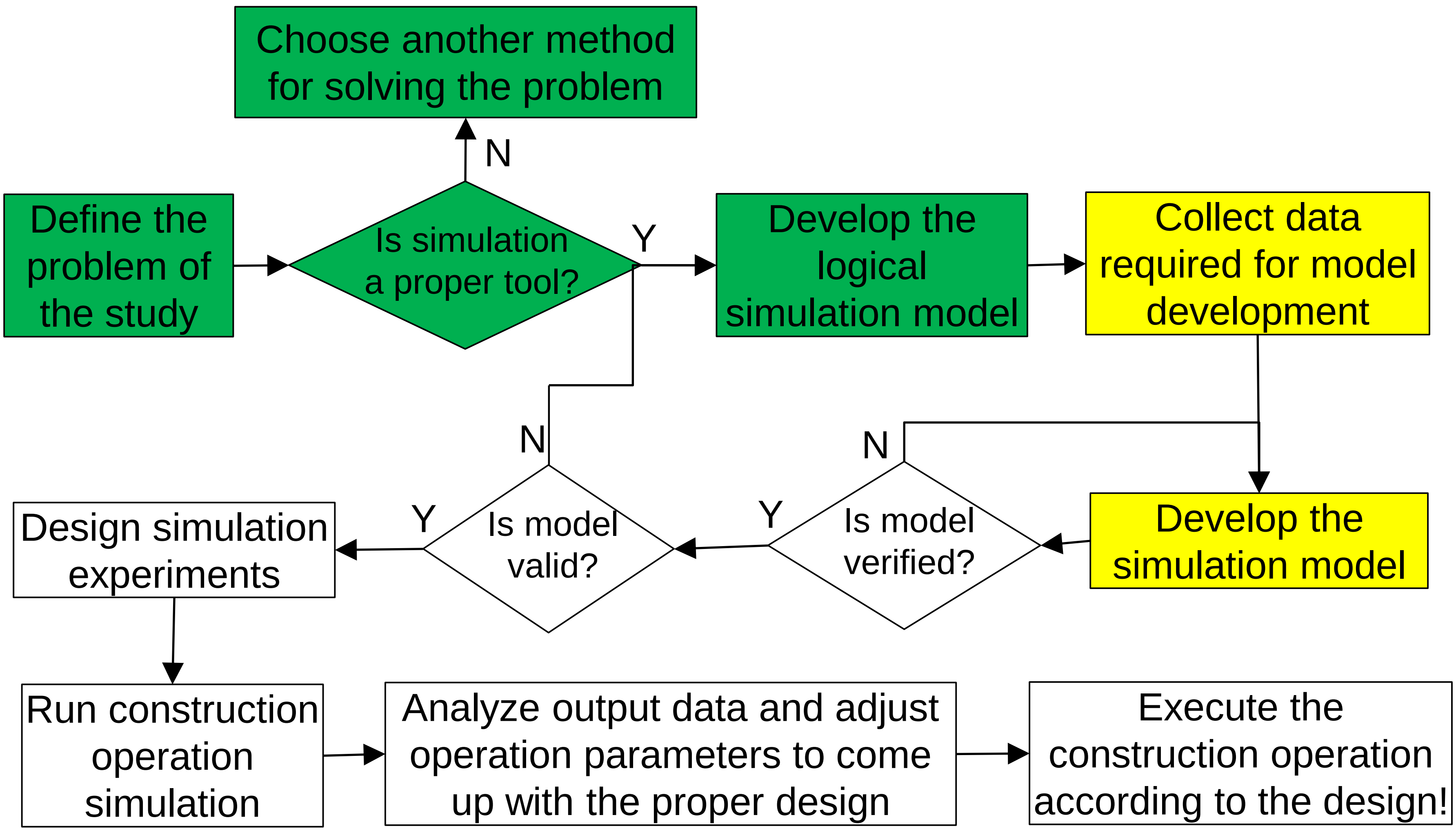
Outline

- Introduction
- Data collection
- Identify a distribution
- Estimate distribution parameters
- Testing for goodness of fit

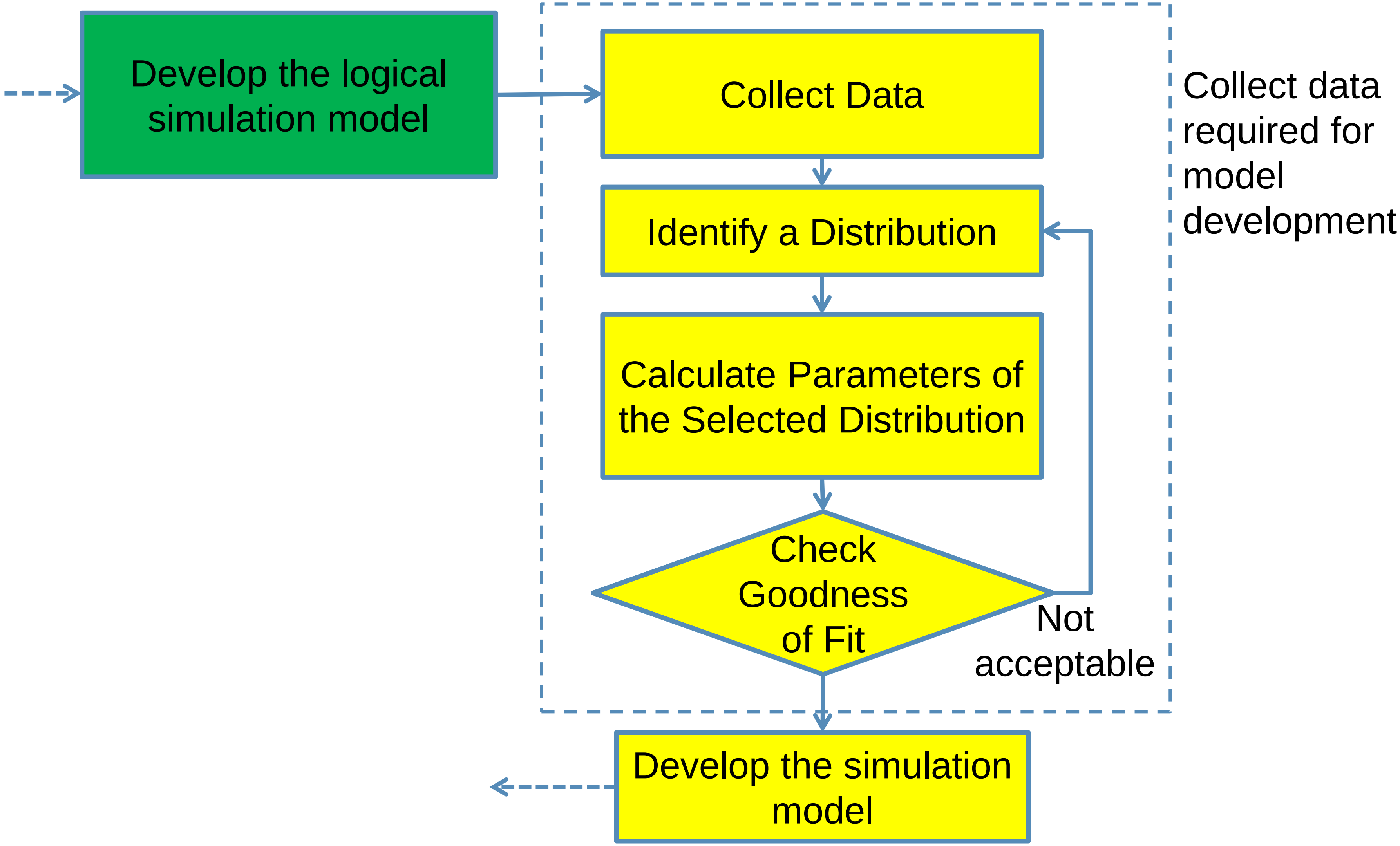
Introduction

- Many aspects of construction operations are stochastic, e.g., activity duration, acceptance/ failure, capacity of equipments, etc.
- These form inputs of our simulation models.
- To be able to properly model construction operations we need to determine these stochastic aspects.
- In general following steps are taken to determine input of simulation models:
  - Collecting data
  - Identifying stochastic distribution to represent data collected
  - Estimate distribution parameter
  - Run goodness of fit for distributions identified and parameters estimated

Introduction



Introduction



Data collection

- There are different ways of data collection:
  - **Using historical data:**
    - Records of past activities collected during the process are valuable sources for estimating activity duration distribution.
  - **Direct observation and sampling:**
    - In cases we face repetitive construction operations and we have not collected historical data, directly observing construction operation and collecting related data required is another way of data collection
  - **Using expert judgement**
    - When we face a new system with no historical records or when there is no historical records available and sampling is costly, using expert judgement is an alternative! We get expert's opinion through interviews and try to converge their opinions. In this case we usually can get some major parameters such as maximum, minimum and most likely values from experts which leads us to simple distributions as inputs. We cannot run goodness of fit at this case since we do not have additional measures to compare these data with!

Data collection

- **Expert judgement (cont'd)**
  - Use of expert judgement in estimating activity durations can result in two types of distributions: *Uniform distribution* and *Triangular distribution*
  - **Uniform distribution** is used for duration estimation when activity experts give a duration range (with a minimum of *a* and a maximum of *b*) for the activity duration. In this case activity duration distribution will be estimated as a uniform distribution:

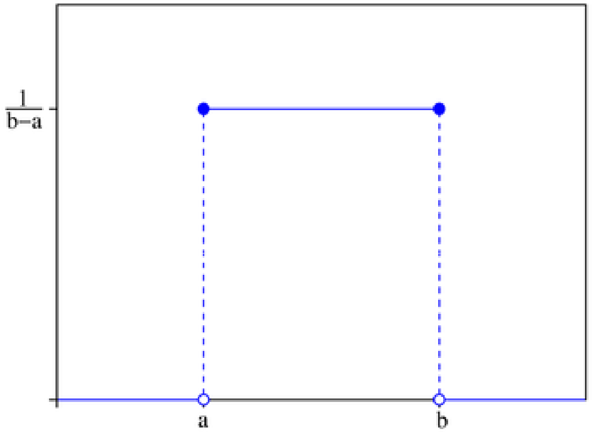
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a, b) \\ 1 & \text{for } x \ge b \end{cases}$$

mean (μ): $\frac{1}{2}(a+b)$

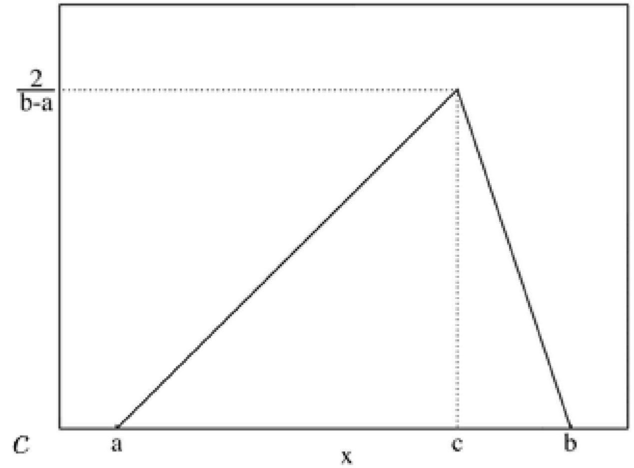
Variance (σ^2): $\frac{1}{12}(b-a)^2$

σ: $\frac{1}{\sqrt{12}}(b-a)$



Data collection

- **Expert judgement (cont'd)**
  - **Triangular distribution** is used for duration estimation when an activity experts give a duration range (with a minimum of *a* and a maximum of *b*) for the activity duration and a most likely value (distribution mode of *c*). In this case activity duration distribution will be estimated as triangular distribution :



$$f(x|a,b,c) = \begin{cases} 0 & \text{for } x < a, \\ \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a \le x \le c, \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c < x \le b, \\ 0 & \text{for } b < x, \end{cases}$$

$$F(x|\ a,\ b,\ \ c) = \begin{cases} 0 & \text{for } x < a, \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{for } a \le x \le c, \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{for } c < x \le b, \\ 1 & \text{for } b < x. \end{cases}$$

Mean(μ): $\dfrac{a+b+c}{3}$

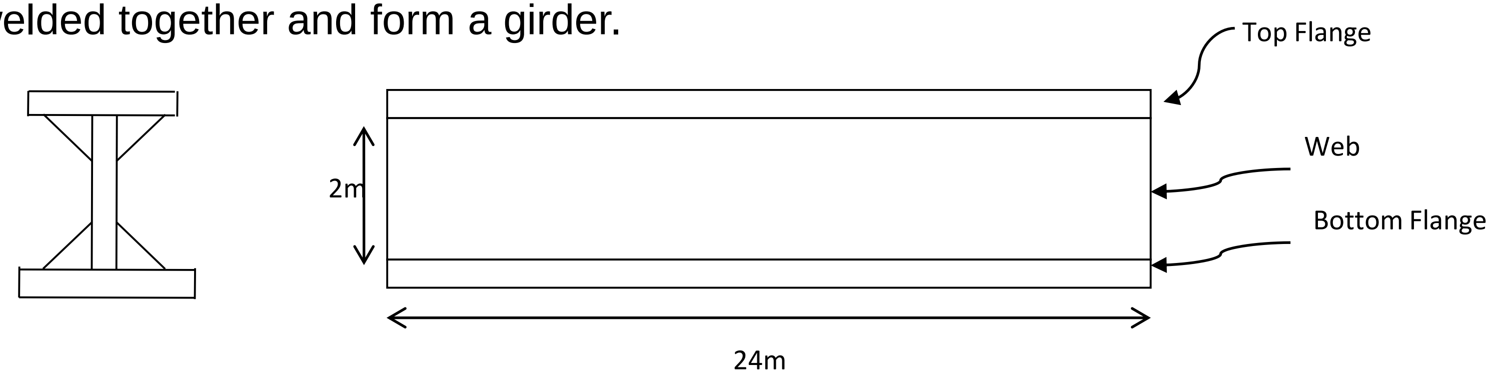
Variance(σ²): $\dfrac{a^2+b^2+c^2-ab-ac-bc}{18}$

σ: $\sqrt{\dfrac{a^2+b^2+c^2-ab-ac-bc}{18}}$

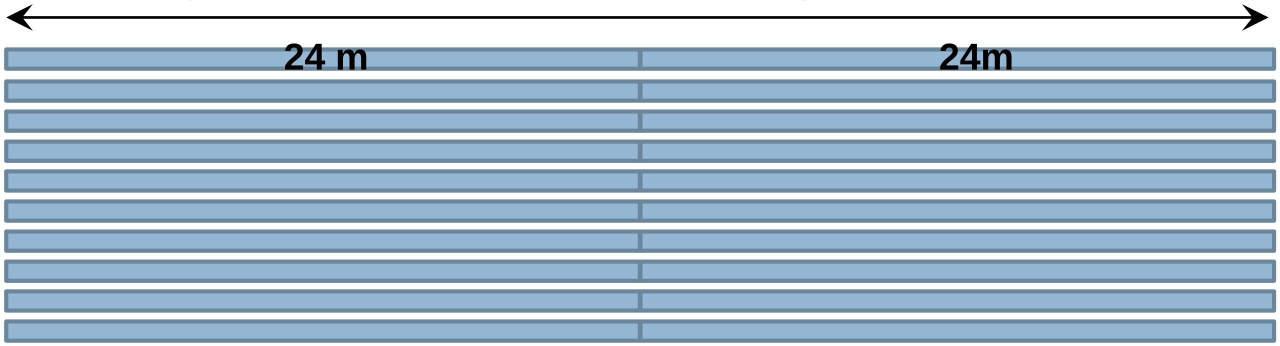
Data collection-Example

Expert judgement-Example:

- We are going to do on-site assemble and install bridge girders on a highway crossing. Bridge has 20 of 24 m girders. Each girder is built from 3 plats which are welded together and form a girder.



- There total number of 10 spans of two-24m girders. To be able to install girders on the top of the bridge column we need to have two girders.



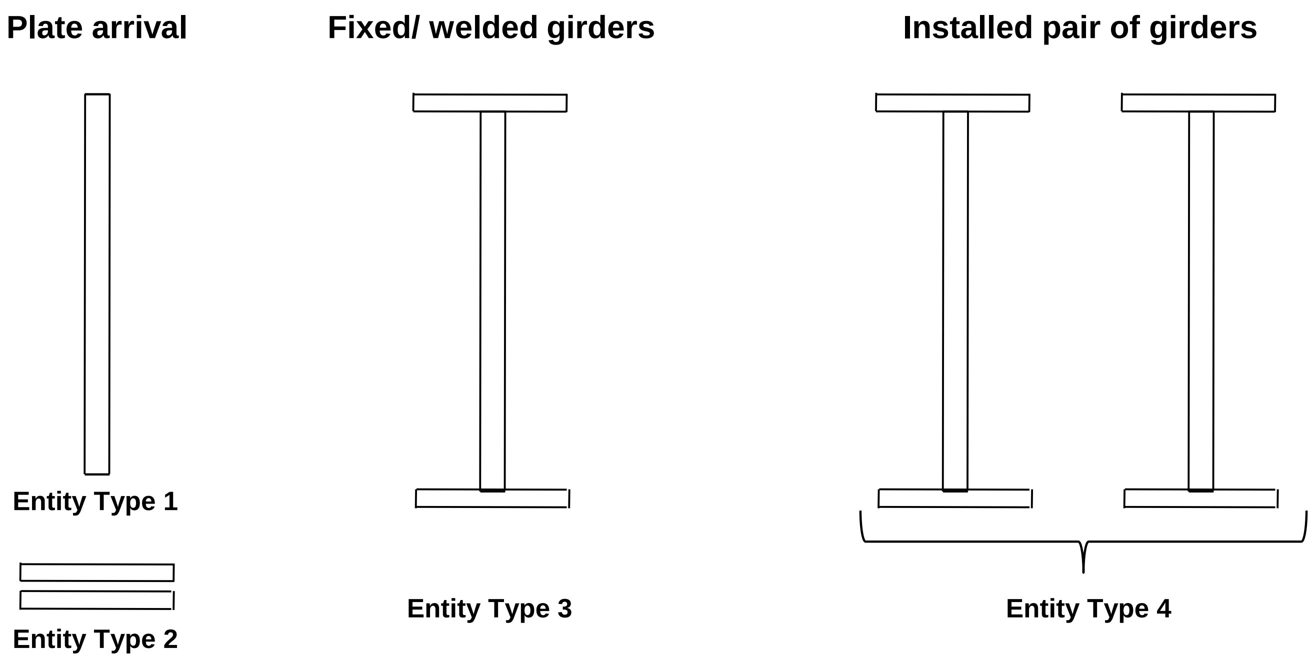
Data collection-Example

- **Expert judgement-example (cont'd):**
  - According to the contract with the sub-contractor, it is going to supply girder plats in separate batches of 5 web plates and 10 flange plates, each of them one batch per day (in average).
  - We need to unload each batch by using 2 iron workers and 2 mobile cranes and put the batches temporary on site storage. According to the superintendent's experience, **unloading** each batch will take a time **between 10 to 20 minutes**. To be able to assemble a girder first we need to fix the plates on their locations. During fixing activity, we need to have two cranes holding the plates and 4 iron workers working. According to the foreman, **fixing** activity will take **at least 30** minutes, but it will not take more than **60 minutes**. However, this will normally take **around 40 minutes. Welding** operation will be performed by 4 welders. It is expected that welding takes **3 to 4 hours**.
  - **Installation** of each pair of girders on a span needs 2 mobile cranes and 2 iron workers and according to the foreman will take about **3 to 5 hours.**
  - With two mobile crane available on site how many days girder installation will take; knowing that construction crew will work 10 hours a day!

Data collection-Example

- **Expert judgement-example (cont’d):**
  - **Entity**: Girder

Do we have girders at the beginning? What should we consider as entity at the beginning?

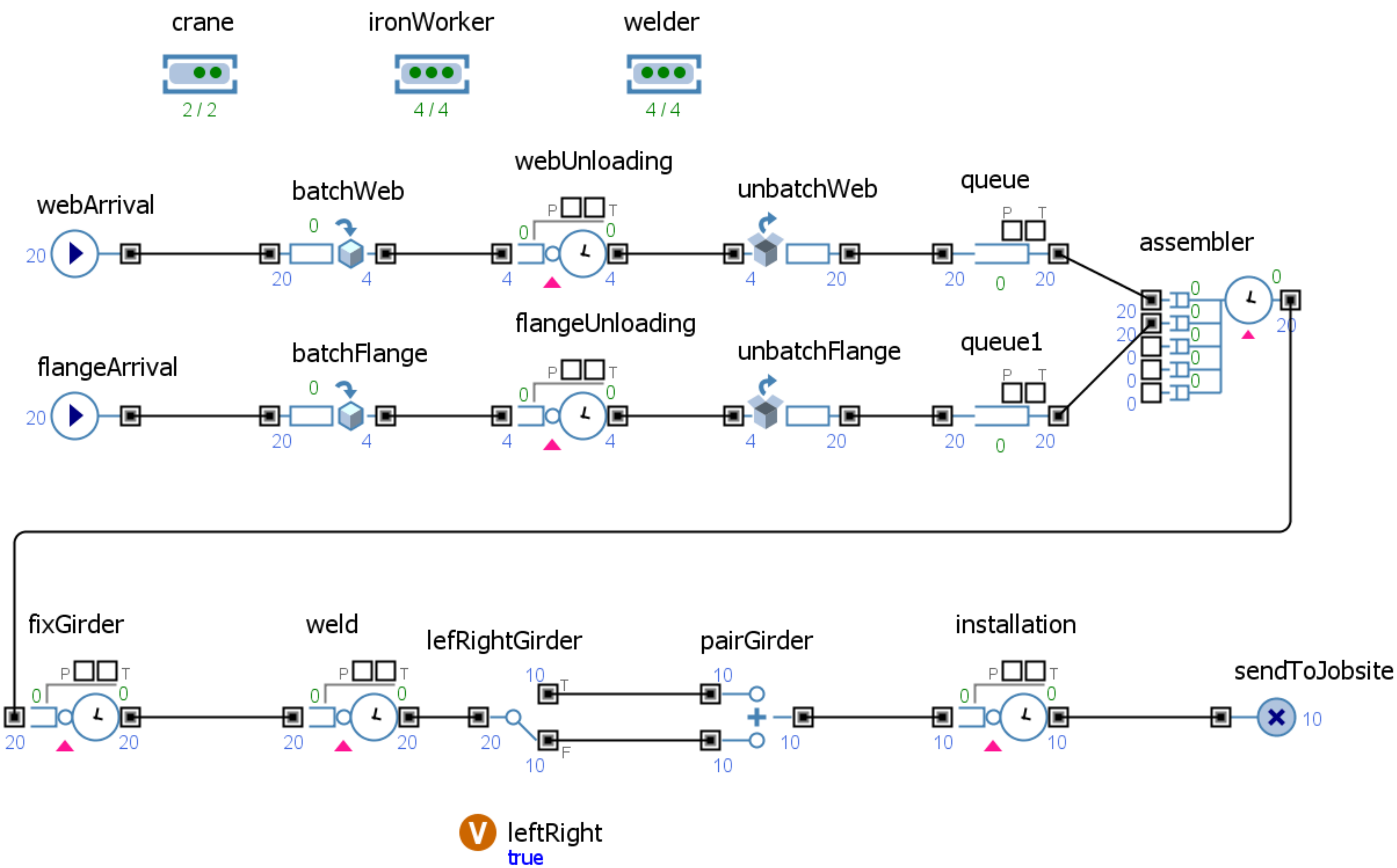


Data collection-Example

- **Expert judgement-example (cont'd):**
  - **Elements:**
    - **Entity**:
      1. We first have Web-plates and Flange-plates as entities
      2. One web plate entity + two flange plates entities form a girder entity
      3. One right side and one left side girder entities for an span entity!
    - **Resource:** 2 mobile cranes, 4 iron workers, and 4 welders
    - **Activity:** Unloading ~ U(10,20), Fixing ~ Triang(30,40,60), Welding ~ U(180, 240), and Installing ~ U(180, 300) (with time unit of minute)
  - **Initial condition:** No material in the system, all resources are idle

Data collection-Example

- Expert judgement-example (cont'd):



Identify a distribution

- In cases we have data collected from an stochastic aspect of a construction operation (using historical data or observation or data sampling) we need to identify a proper distribution that data can fit into it.
- Our main method for identifying the distribution is to present data on a **histogram**, and then see if histogram's shape is close to probability density function curve of any distribution! Following steps are taken for drawing a histogram:

  1. Divide the range of the data into intervals. (Intervals are usually of equal width)

     - One concern to divide the range data into intervals is to set an adequate number of intervals; too many or too few intervals are misleading. A guideline point is to select the number of intervals close to square root of the number of data collected, e.g., for 100 data items divide them into 10 intervals.

     - Another concern is number of data items collected. We need a minimum number of data items collected to be able to unbiasedly identify the distribution. A guideline for the minimum number of data collected is 25!

Identify a distribution

2. Label the horizontal axis to conform to the intervals selected.
3. Find the frequency of occurrences within each interval and label the vertical axis so that the total occurrences can be plotted for each interval.
4. Plot the frequencies on the vertical axis.

Example: Following historical data is collected from truck trip time to the dumping site:

Trip time (minute)	Trip time (minute)	Trip time (minute)	Trip time (minute)
26.7	31.7	36.3	29.6
23.9	28.4	32.1	27
27.6	23.7	35.6	36.6
25.7	24.3	29.6	30.1
37.2	18.8	19.6	21.3
26.1	28	34.2	26.1
34.1	25.4	31.5	32
20	16.6	32.9	26.6
27.7	37	27.4	24.9
26.4	30.8	29.2	26.9
29.6	34.7	30.1	23.9
29.1	28.3	35.8	23

Identify a distribution

Example (cont'd):

1- Total number of historical data are 48; its square root is close to 7; we are going to have 7 intervals!
Minimum value is **16.6** And maximum value is **37.2** so total intervals length will be **20.6** minutes and every interval is **2.94**. Border values for our intervals are:

16.6	19.54	22.48	25.42	28.36	31.3	34.24	37.2

2- Middle value of every interval is going to be representative value of each interval and is going to be presented on the horizontal axis of the histogram:

Interval #	1	2	3	4	5	6	7
Rep. Value	18.07	21.01	23.95	26.89	29.83	32.77	35.72

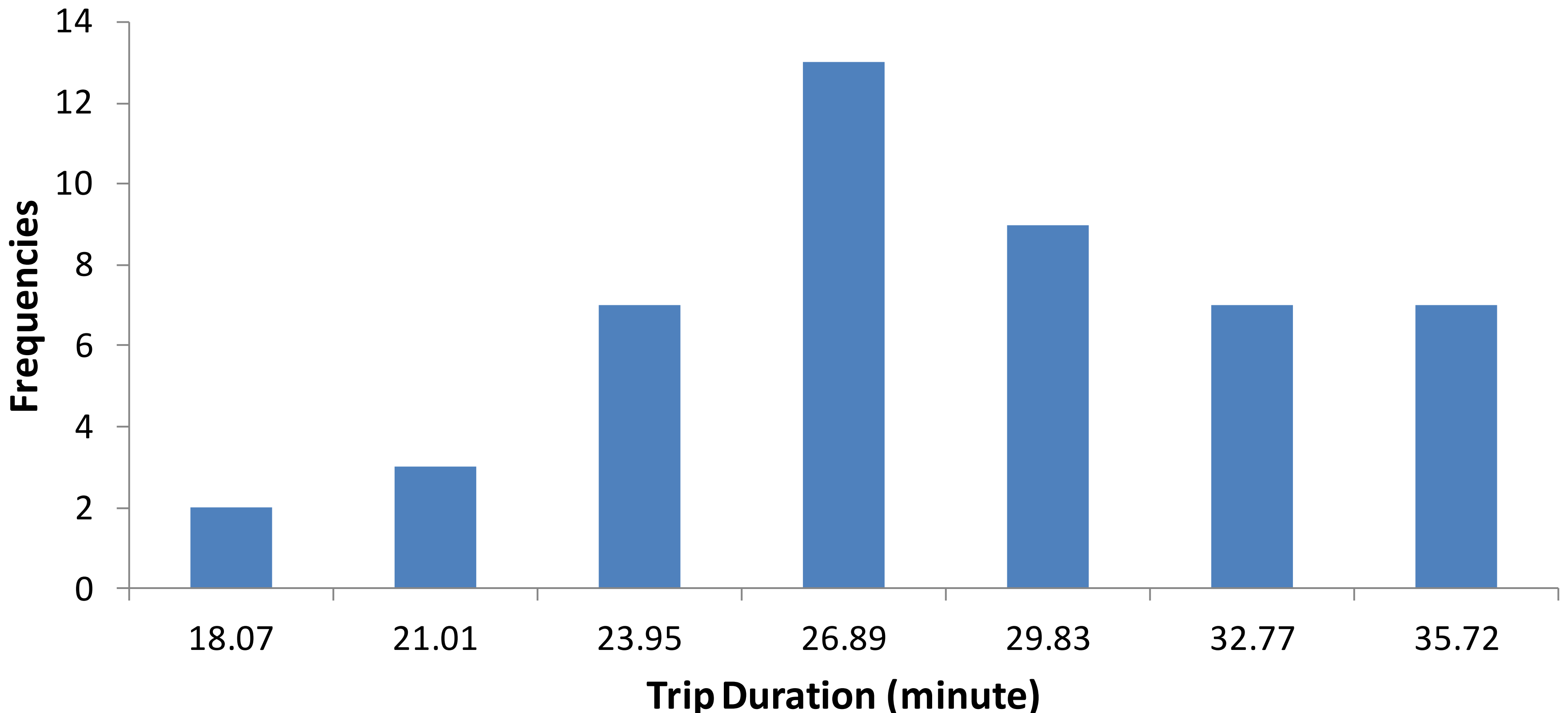
3- Frequency of occurrence within in each interval is:

Interval #	1	2	3	4	5	6	7
Freq.	2	3	7	13	9	7	7

Identify a distribution

Example (cont'd):

4- Plotting the histogram



What distribution this histogram might represent?

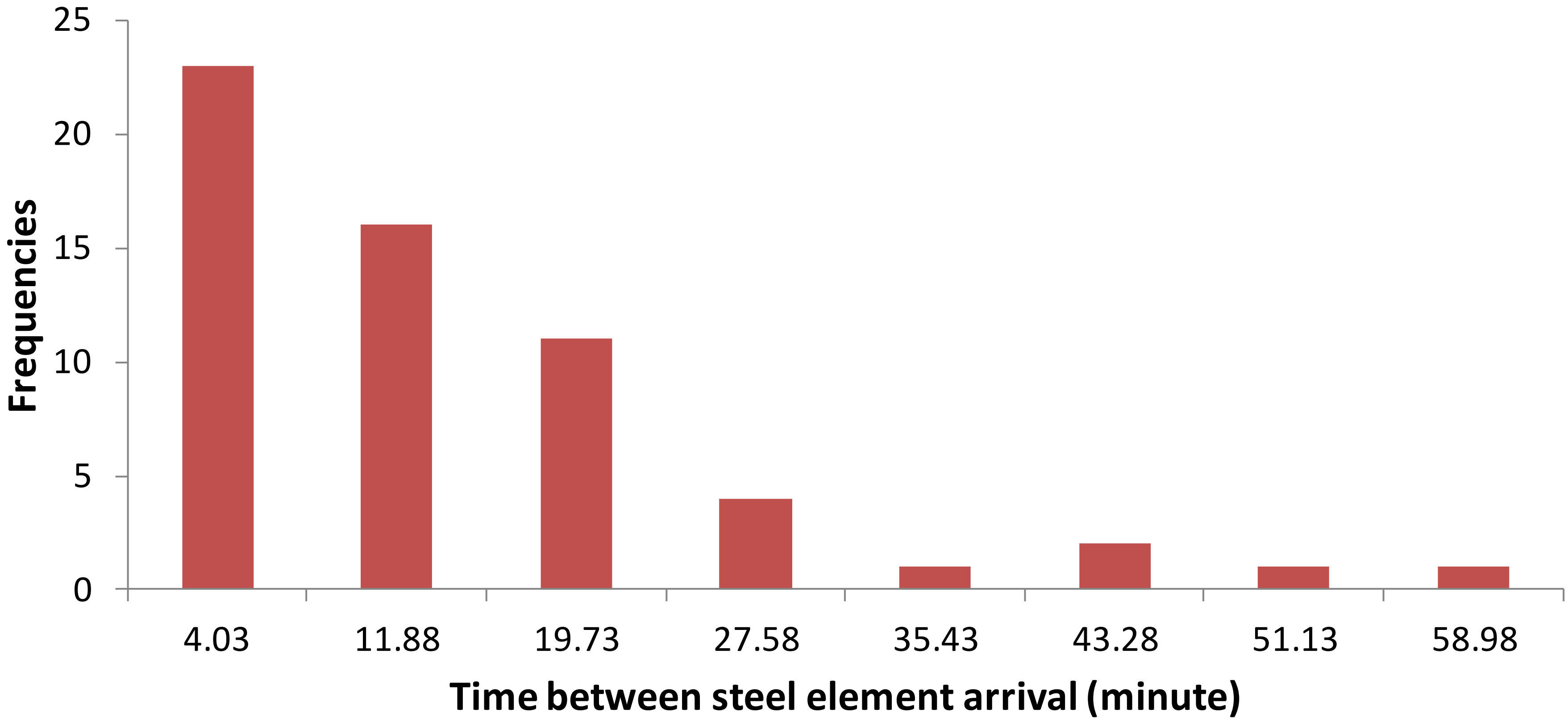
Identify a distribution

Example: Following historical data collected for time between arrivals of steel elements arrive to the construction site:

Time between arrival (minute)	Time between arrival (minute)	Time between arrival (minute)
17.4	13.4	11.6
45.1	24.7	35.3
6.4	6.6	1.8
11	16.3	61
3.3	17.3	1.9
18.1	6.6	2.6
0.1	0.3	14.2
7.9	16.8	20.2
19.1	13.8	17.9
20.9	6.3	3.8
10.8	8.4	49
1.5	62.9	15.6
3.1	8.3	20.4
1.2	9.5	28.1
6.5	11.9	6.3
0.2	12.8	13.4
15.2	0.9	39.8
0.8	23.7	29.1
5.6	8.5	5.2
2.7	8.9	20.6

Identify a distribution

Example (cont'd):



What distribution this histogram might represent?

Estimating distribution parameters

Maximum Likelihood Estimation (MLE):

The most common method for estimating a distribution parameter is MLE.

In words we can define likelihood function of our distribution parameter (Ɵ) as:

like(Ɵ)= Probability of observing the given data as a function of Ɵ

= f(x1, x2, ..., xn| Ɵ)

= f(x1| Ɵ) f(x2| Ɵ) ... f(xn| Ɵ) = ∏ f(xi| Ɵ)

Max like(Ɵ)= Max ∏ f(xi| Ɵ)

Note: Most of the time this product takes the parameter into exponent and we would rather to use the maximization of logarithm of the product function or the exponent.

Max like(Ɵ)=> Max l(Ɵ)= Max log(∏ f(xi| Ɵ))

Example: Use of MLE for estimating Normal distribution parameters

$$f(x1, x2, \ldots, xn|\mu) = \prod_{i}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{xi-\mu}{\sigma})^2}$$

Estimating distribution parameters

Example (cont'd):

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \times e^{\underbrace{-\frac{1}{2\sigma^2}\overbrace{\sum_i^n (xi-\mu)^2}^{\textbf{minimize}}}_{\textbf{maximize}}}$$

$$max\ like(\mu) = min\sum_i^n (xi-\mu)^2 = \min\left(n\mu^2 + \sum_i^n xi^2 - 2\mu\sum_i^n xi\right)$$

$$\frac{\partial\left(n\mu^2 + \sum_i^n xi^2 - 2\mu\sum_i^n xi\right)}{\partial\mu} = 2n\mu - 2\sum_i^n xi = 0$$

$$=> \quad \mu_e = \bar{X} = \frac{\sum_i^n xi}{n}$$

Estimating distribution parameters

Example (cont'd): for estimating σ we have:

$$like(\sigma) = f(x1, x2, \dots, xn|\sigma) = \prod_{i}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{xi-\mu}{\sigma})^2}$$

$$= (\frac{1}{\sigma\sqrt{2\pi}})^n \times e^{-\frac{1}{2\sigma^2}\sum_i^n (xi-\mu)^2}$$

$$l(\sigma) = n\, ln(\frac{1}{\sigma\sqrt{2\pi}}) - \frac{1}{2\sigma^2}\sum_{i}^{n}(xi-\mu)^2$$

$$l(\sigma) = -n\, ln\, \sigma - n\, ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_{i}^{n}(xi-\mu)^2$$

$$\frac{\partial l(\sigma)}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i}^{n}(xi-\mu)^2 = 0 \quad => \quad \sigma_e = \sqrt{\frac{\sum_i^n (xi-\mu_e)^2}{n}}$$

Estimating distribution parameters

Example (cont'd):

Note: **σ_e** here is a bias estimator for **σ** or standard deviation of the normal distribution and is called the sample estimator. It is proved that an unbiased estimator for **σ** is:

$$\sigma_e = S = \sqrt{\frac{\sum_i^n (xi - \mu_e)^2}{n-1}}$$

Estimating distribution parameters

Suggested estimators for some distributions used in simulation

Distribution	*Parameter(s)*	*Suggested Estimator(s)*
Poisson	α	$\hat{\alpha} = \bar{X}$
Exponential	λ	$\hat{\lambda} = \dfrac{1}{\bar{X}}$
Normal	μ, σ^2	$\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = S^2$ (unbiased)
Beta	β_1, β_2	$\Psi(\hat{\beta}_1) + \Psi(\hat{\beta}_1 - \hat{\beta}_2) = \ln(G_1)$ $\Psi(\hat{\beta}_2) + \Psi(\hat{\beta}_1 - \hat{\beta}_2) = \ln(G_2)$ where Ψ is the digamma function, $G_1 = \left(\prod_{i=1}^{n} X_i\right)^{1/n}$ and $G_2 = \left(\prod_{i=1}^{n} (1 - X_i)\right)^{1/n}$

Testing for goodness of fit

- Having parameterized a distribution, one should check for the goodness of fit by comparing the fitted distribution to the empirical distribution and assessing the quality of the fit obtained.
- Usually one would perform the goodness of fit test by using:
  - Visual assessment
  - Chi-Square test
  - Kolmogorov-Smirnov (K-S) test

Testing for goodness of fit

- **Visual assessment**
  - Basically the method is simply to plot both the empirical and fitted CDFs on one plot and compare how well the fitted CDF tracks the empirical one.
  - Alternatively one can compare the how well the shape of the sample histogram compares to that of the theoretical PDF.
  - When the CDF is available, it is always better to compare to it because, a histogram can be easily distorted, by changing number of intervals can attain practically different shapes.
  - Although the method does not bear much statistical or mathematical weight, it remains as effective as any other test.

Testing for goodness of fit

- **Visual assessment**
  - **Example**:

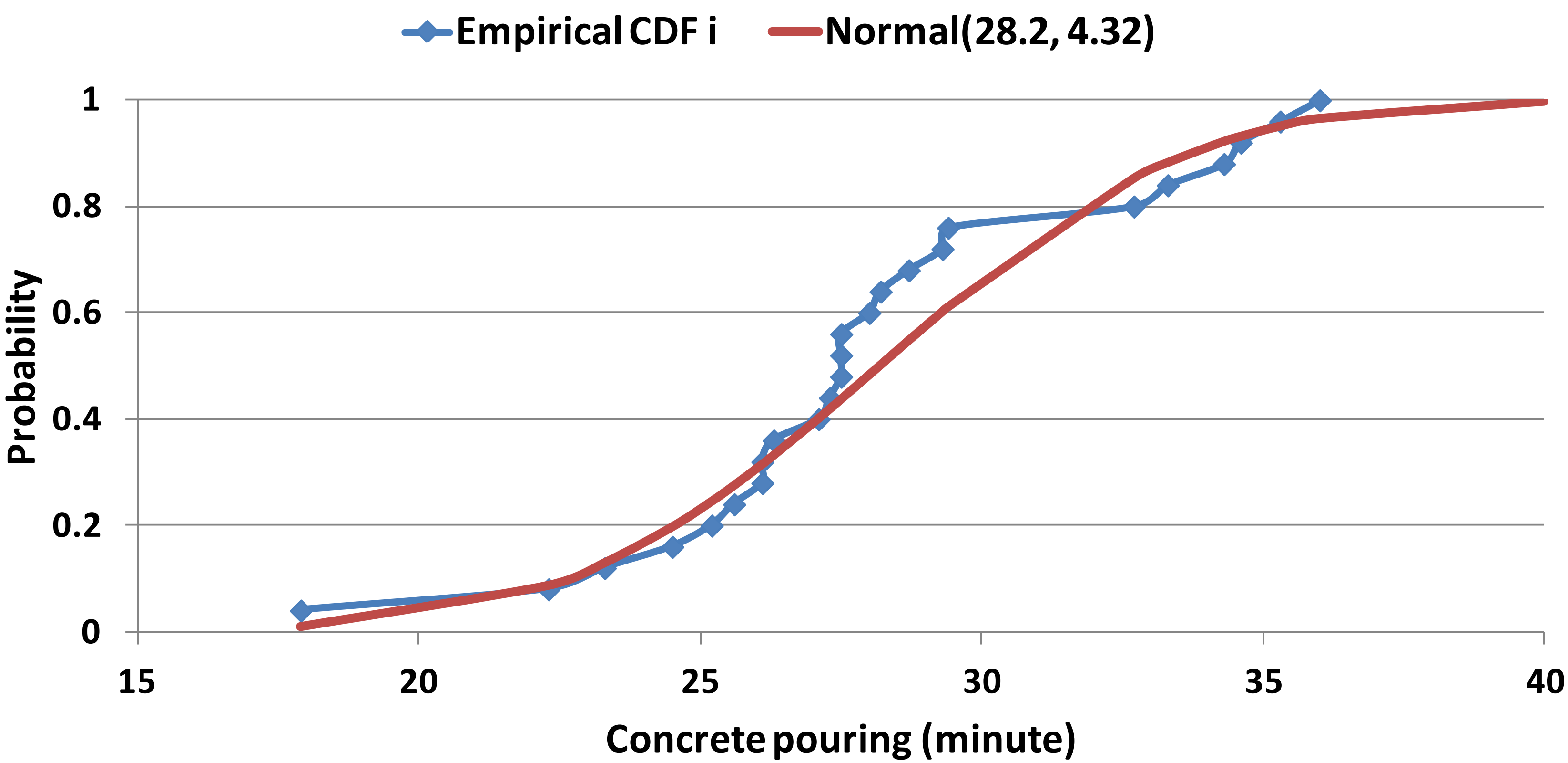
    Normal distribution has been identified for concrete pouring activity with mean of 28.2 and standard deviation of 4.32 minute/m3.

    In our visual assessment we have:

i	Concrete pouring (minute/m3)	Empirical CDF i	Normal (28.2, 4.32)
1	17.9	0.04	0.009
2	22.3	0.08	0.086
3	23.3	0.12	0.128
4	24.5	0.16	0.196
5	25.2	0.2	0.244
6	25.6	0.24	0.274
7	26.1	0.28	0.313
8	26.1	0.32	0.313
9	26.3	0.36	0.330
10	27.1	0.4	0.400
11	27.3	0.44	0.417
12	27.5	0.48	0.436
13	27.5	0.52	0.436
14	27.5	0.56	0.436
15	28	0.6	0.482
16	28.2	0.64	0.500
17	28.7	0.68	0.546
18	29.3	0.72	0.600
19	29.4	0.76	0.609
20	32.7	0.8	0.851
21	33.3	0.84	0.881
22	34.3	0.88	0.921
23	34.6	0.92	0.931
24	35.3	0.96	0.950
25	36	1	0.965

Testing for goodness of fit

- **Visual assessment**
  - **Example (cont'd)**:



Testing for goodness of fit

- **Chi-square test:**
  - Karl Pearson's Chi-Square test is based on the measurement of the discrepancy between the histogram of the sample and the fitted probability density function.
  - This test requires large sample sizes.
  - It tests the observations Xi (i=1, 2, ...,n) following a particular distribution f(x) where we computed distribution parameters using the MLE approach:

1. The test procedure begins by arranging the n observations into a set of k class intervals or cells (C1, C2, .., Ck).

   a. Class intervals should be of equal weights (expected probabilities) of 1/k; so that F(Ci) = 1/k = P.

   b. Pick number of class intervals close to square root of number of observations.

   c. Minimum expected number of observations in each class should not be less than 5

Testing for goodness of fit

- **Chi-square test:**

2. Count the number of observation X_i (i=1,2,…,n) within each cell of C_j *(j=1,2, …, K) noted as Nj (j=1, 2, …, K)*

3. Calculate Chi-Square statistic χ^2 as:

$$\chi^2 = \sum_{j=1}^{k} \frac{(Nj - np)^2}{np}$$

Where n is the total number of observation, N_j is the respective number of observation in cell j and $p=1/k$

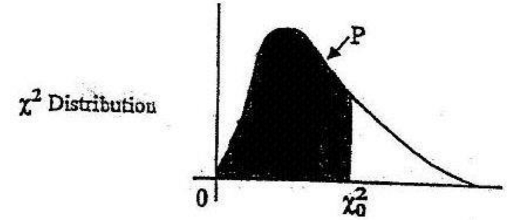
4. Compare the value of χ^2 obtained to that from the Chi-Square tables. Reject the distribution if

$$\chi^2_{\text{computed}} > \chi^2_{1-\alpha,\,(\text{k-s-1})}$$

Where; α is level of error (type one) accepted, 1- α or P value is level of confidence, k is number of classes, s is number of estimated parameters and *k-s-1* is the degree of freedom.

Testing for goodness of fit

- Chi-square test:
  - Chi-square table



The table below gives the value x_0^2 for which $P[x^2 < x_0^2] = P$ for a given number of degrees of freedom and a given value of P.

Degrees of Freedom	Values of P				
	0.900	0.950	0.975	0.990	0.995
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267
17	24.769	27.587	30.191	33.409	35.718
18	25.989	28.869	31.526	34.805	37.156
19	27.204	30.144	32.852	36.191	38.582
20	28.412	31.410	34.170	37.566	39.997

Testing for goodness of fit

- **Chi-square test:**

  - **Example:** Following historical data of truck trip duration to the dumping site is identified to have a family distribution of normal. Estimate distribution parameters and run a chi-square test to test the goodness of the fit.

Trip time (minute)	Trip time (minute)	Trip time (minute)	Trip time (minute)
26.7	31.7	36.3	29.6
23.9	28.4	32.1	27
27.6	23.7	35.6	36.6
25.7	24.3	29.6	30.1
37.2	18.8	19.6	21.3
26.1	28	34.2	26.1
34.1	25.4	31.5	32
20	16.6	32.9	26.6
27.7	37	27.4	24.9
26.4	30.8	29.2	26.9
29.6	34.7	30.1	23.9
29.1	28.3	35.8	23

Testing for goodness of fit

- **Chi-square test:**
  - **Example (cont'd):**

$$\mu_e = \bar{X} = \frac{\sum_i^n xi}{n} = 28.42 \qquad \sigma_e = \sqrt{\frac{\sum_i^n (xi - \mu_e)^2}{n}} = 4.94$$

1. Sample Number = 48 => Num of interval = 7 And P=1/7= 0.143

i	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
CDF	0	0.143	0.286	0.429	0.571	0.714	0.857	1
Ci	- ∞	23.1	25.6	27.5	29.3	31.2	33.7	+ ∞

Testing for goodness of fit

- **Chi-square test:**
  - **Example (cont’d):**

2. Number of observations for each cell

i	1	2	3	4	5	6	7
Ni	6	6	9	7	6	5	9

3. Chi-square value is:

$$\chi^2 = \sum_{j=1}^{k} \frac{(Nj - np)^2}{np} = 2.167$$

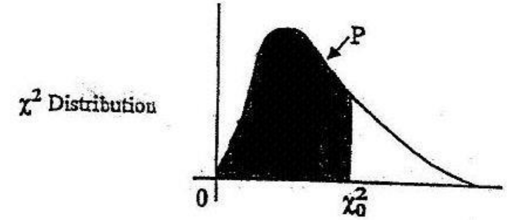
4. To compare calculated value with chi-square with 4 (=7-2-1) degree of freedom and confidence level of 95% (read from table) we have:

$$\chi^2_{\text{computed}} = 2.167 < \chi^2_{95\%,\,4} = 9.488$$

So, goodness of distribution is accepted!

Testing for goodness of fit

- **Chi-square test:**
  - **Example (cont'd):**



The table below gives the value x_0^2 for which $P[x^2 < x_0^2] = P$ for a given number of degrees of freedom and a given value of P.

Degrees of Freedom	Values of P				
	0.900	0.950	0.975	0.990	0.995
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267
17	24.769	27.587	30.191	33.409	35.718
18	25.989	28.869	31.526	34.805	37.156
19	27.204	30.144	32.852	36.191	38.582
20	28.412	31.410	34.170	37.566	39.997

After class practice

Following data are collected from time between equipment breakdown on site which is suggested to have exponential distribution. Estimate the distribution's parameter and use chi-square test with the error level of 5% to test its goodness of fit!

Time between break down (hour)	Time between break down (hour)	Time between break down (hour)
55	60	6
103	12	107
69	24	31
53	32	0
188	230	94
284	0	36
27	87	108
91	78	27
36	43	98
18	148	41
265	109	362
41	160	119
64	119	7
36	62	133
178	36	73
79	27	103

Testing for goodness of fit

- **Kolmogorov-Smirnov Test**
  - The test is based on the measurement of the discrepancy between the empirical CDF and fitted CDF.

the max positive deviation given by :

$$D_n^+ = \max\left\{ \,|\, F(X_i)_{empirical} - F(X_i)_{theoretical} \,|\right\} = \max\left\{ \left| \frac{i}{n} - F(X_i) \right| \right\} \;\; for\ i = 1,\ \ldots,\ n$$

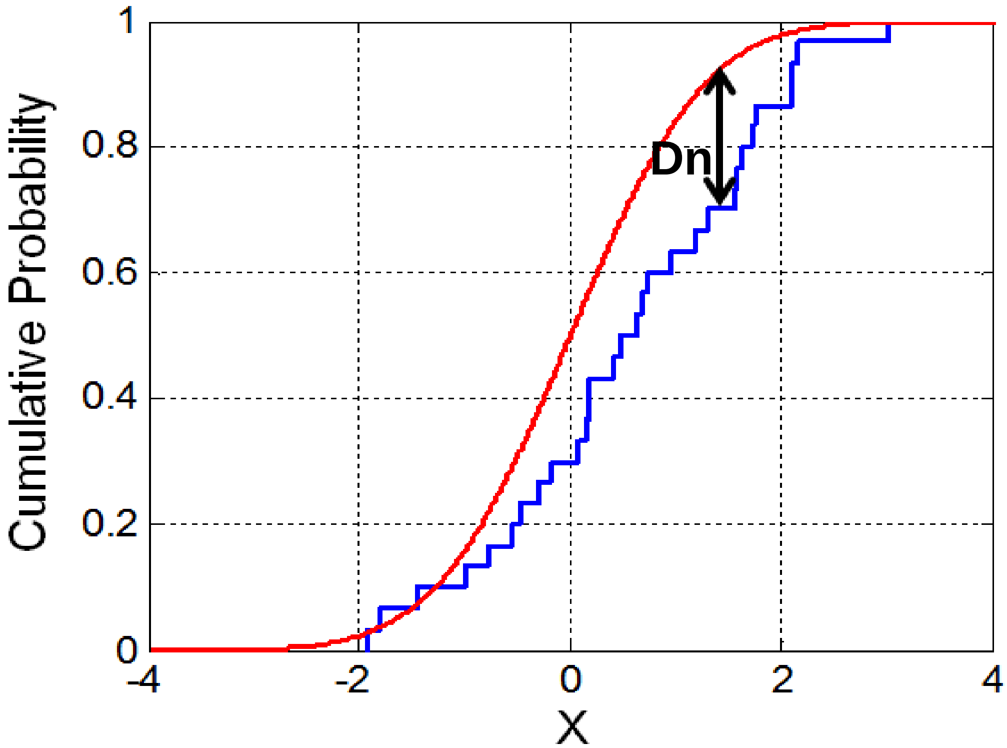
the max negative deviation given by :

$$D_n^- = \max\left\{ \,|\, F(X_i)_{theoretical} - F(X_{i-1})_{empirical} \,|\right\} = \max\left\{ \left| F(X_i) - \frac{i-1}{n} \right| \right\} \;\; for\ i = 1\ldots\ n$$

the test statistic would be $D_n = \max\left\{ D_n^+, D_n^- \right\}$

Testing for goodness of fit

- **Kolmogorov-Smirnov Test**
  - **Illustration of the Kolmogorov-Smirnov statistic.**



Red line is Theoretical CDF, blue line is an Empirical CDF, and the black arrow is the K-S statistic.

Testing for goodness of fit

- **Kolmogorov-Smirnov Test**
  - The test would be to compute D_n as described above and to compare the value with that from a table of critical K-S values at the appropriate confidence and degrees of freedom.
  - When the computed D is larger than the theoretical one the distribution is rejected.
  - The Kolmogorov-Smimov test is particularly useful when sample sizes are small and when no parameters have been estimated from the data.

Testing for goodness of fit

- **Kolmogorov-Smirnov Test**
  - **K-S table**

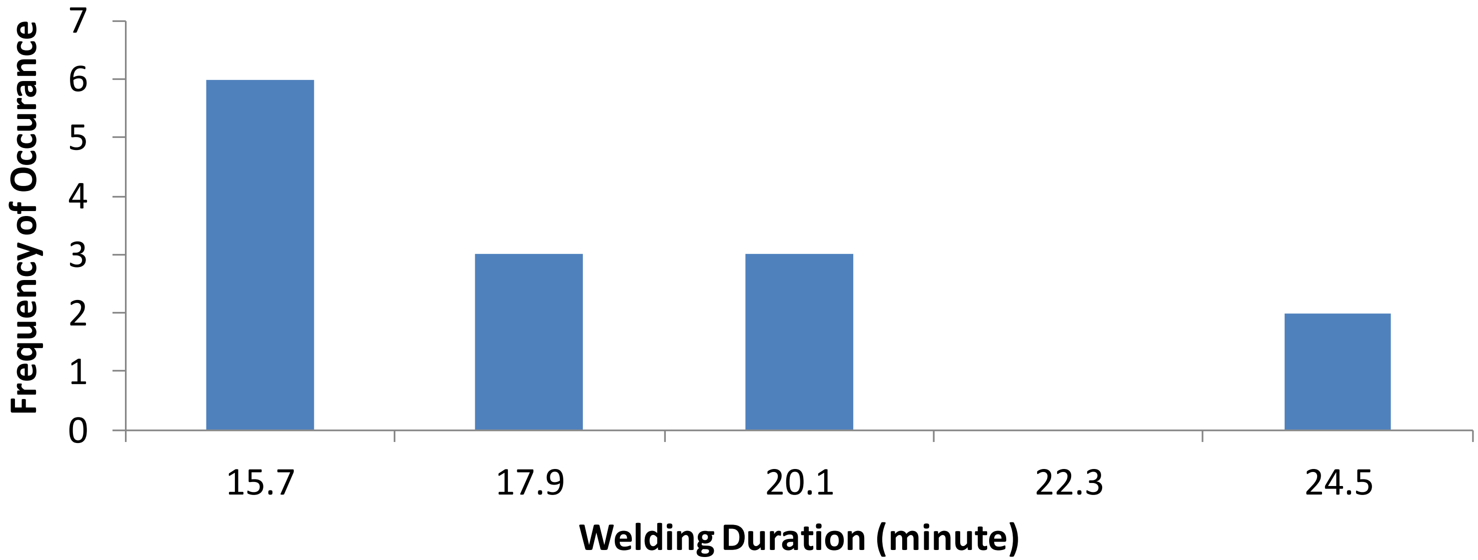
Degrees of Freedom (*N*)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27
Over 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Testing for goodness of fit

- **Kolmogorov-Smirnov Test**
  - **Example:** 15 duration samples of a welding task is given as in below:

Welding duration (minute)	14.6	15.6	15.8	16.3	16.6	16.6	16.9	18.5	17	20.1	20.4	21.1	23.5	25.3	25.6

- Presenting occurrence frequencies of the sample in 5 equal interval we have



Testing for goodness of fit

- **Kolmogorov-Smirnov Test**
  - **Example (cont'd):**
  - Based on the graph shape and according to the expert judgment it is suggested that data duration might have a triangular distribution with minimum and most likely value of minimum of the sample minus 1 and maximum value of maximum of the sample plus 1.

    Welding duration ~ Triangular[13.6, 13.6, 26.6]

Testing for goodness of fit

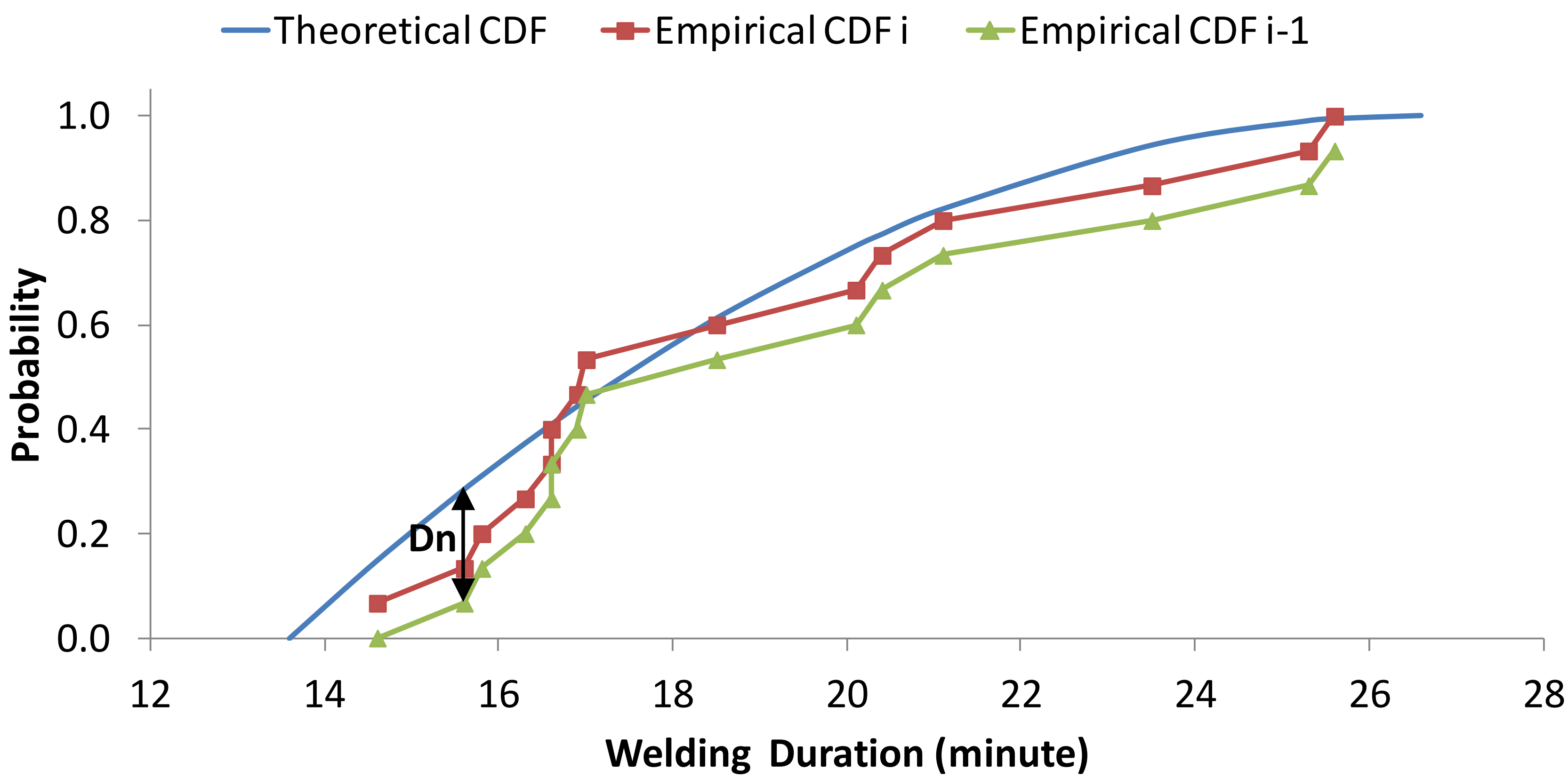
- Kolmogorov-Smirnov Test
  - Example (cont'd):

i	Welding duration (minute)	Emperical Fxi	Emperical Fxi-1	Theoritical Fxi	D+	D-
1	14.6	0.07	0.00	0.15	0.08	0.15
2	15.6	0.13	0.07	0.28	0.15	0.22
3	15.8	0.20	0.13	0.31	0.11	0.18
4	16.3	0.27	0.20	0.37	0.11	0.17
5	16.6	0.33	0.27	0.41	0.07	0.14
6	16.6	0.40	0.33	0.41	0.01	0.07
7	16.9	0.47	0.40	0.44	0.02	0.04
9	17	0.60	0.47	0.45	0.15	0.01
8	18.5	0.53	0.60	0.61	0.08	0.01
10	20.1	0.67	0.53	0.75	0.08	0.22
11	20.4	0.73	0.67	0.77	0.04	0.11
12	21.1	0.80	0.73	0.82	0.02	0.09
13	23.5	0.87	0.80	0.94	0.08	0.14
14	25.3	0.93	0.87	0.99	0.06	0.12
15	25.6	1.00	0.93	0.99	0.01	0.06
					Max Dn	**0.22**

Testing for goodness of fit

- Kolmogorov-Smirnov Test
  - **Example (cont'd):**

K-S test illustration



Testing for goodness of fit

- **Kolmogorov-Smirnov Test**
  - **Example (cont'd):**

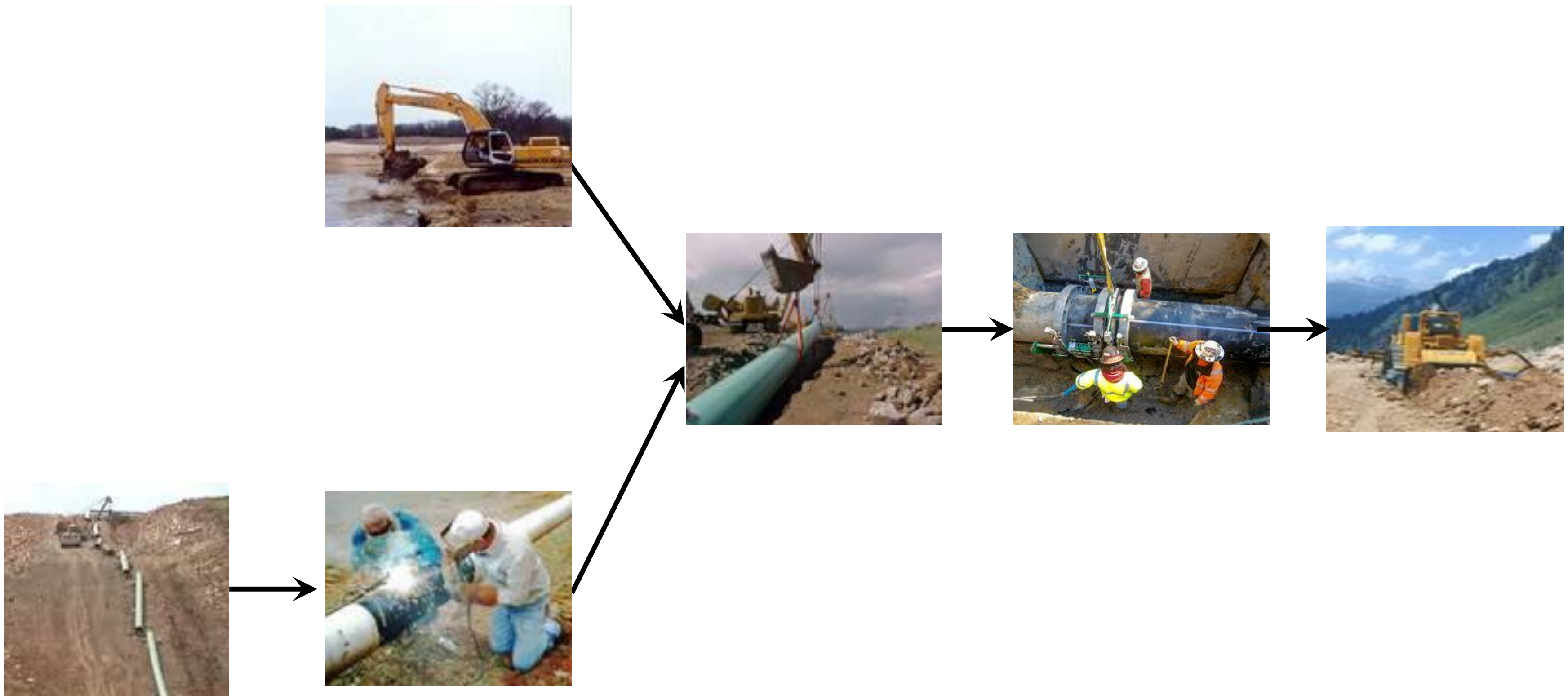
    0.22 < 0.338 => We accept the distribution

Degrees of Freedom (N)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27
Over 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Pipeline construction example

- This example is a simplified real case study!
- We are going to help a 30 km of 40 inch pipeline construction project in the south part of the country by simulating the whole operation using AnyLogic for balancing resources required in this project.
- The main steps of the operation involve:
  1) Receiving pipes from fabrication shops
  2) Welding pipes
  3) Digging a ditches
  4) Lowering pipes into the ditch
  5) Coupling pipes in the ditch
  6) Back filling the dirt to the ditch

Pipeline construction example



Pipeline construction example

- 10 meter length pipes required in the project are fabricated by 3 different industrial pipe fabrication companies in the country. Fabricated pipes are shipped to the construction site in a pack of 3 pipes. It is expected that every week each fabrication shop can send 10 packs of pipes to the construction site.
- A shovel arm is used to unload pipes and string them beside the ditch. According to the foreman this will take 20 to 30 minutes to unload and string pipes beside the ditch.
- Shovel is also used to dig the ditch. According to the foreman's past experience shovel can dig every meter of the ditch in 10 to 15 minutes.

Pipeline construction example

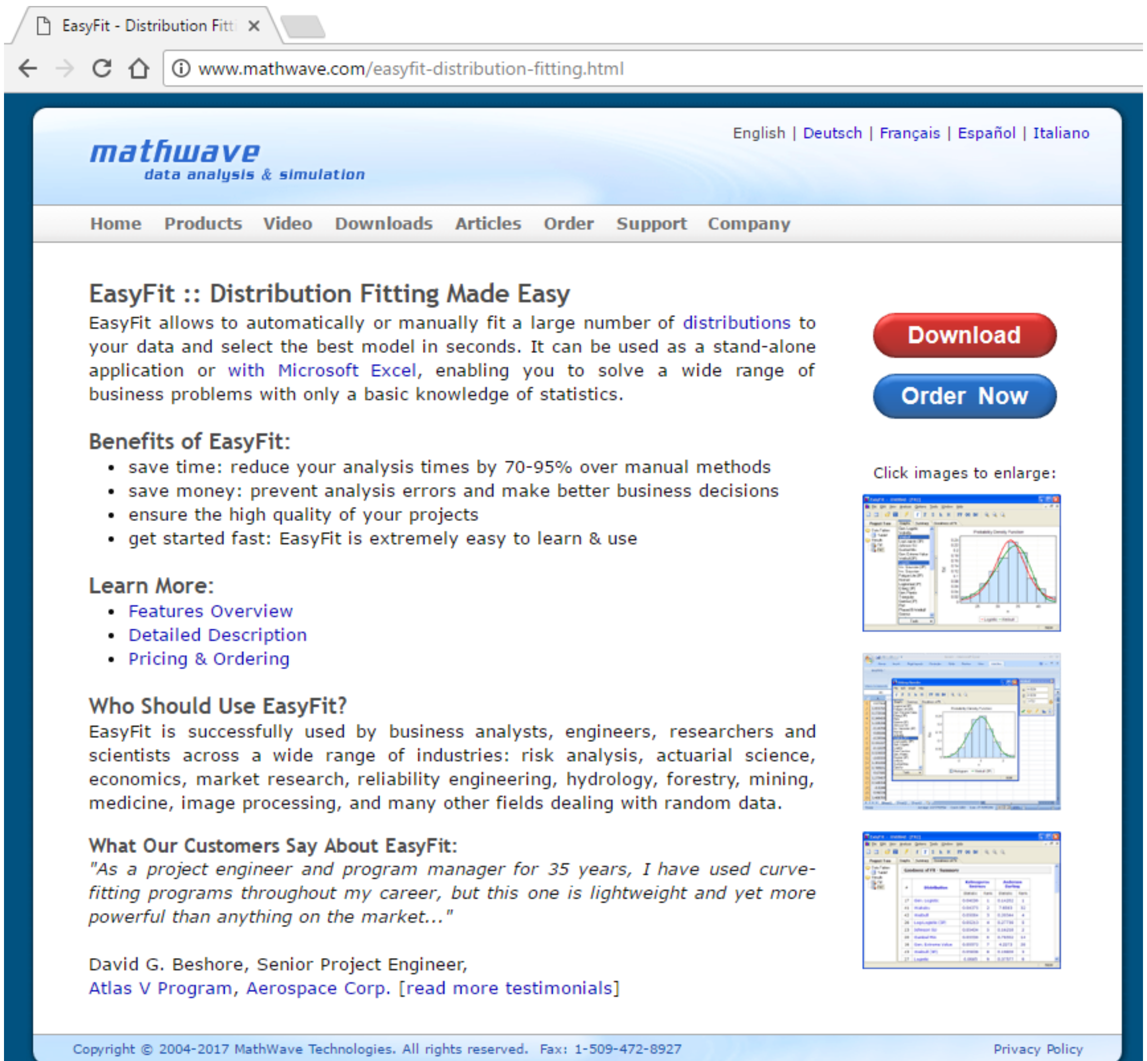
- A welding crew first welds 10 pipes together when they are sitting beside the ditch.
- Then set of 10 welded pipes are lowered into the ditch using the shovel arm with the coordination of welding crew. Lowering in 10 welded pipes into the ditch usually takes 2 to 3 hours.
- One set of 10-welded pipe on the ditch is then coupled with another set of 10-welded pipes and sealed by a welding crew.
- Sample duration time for welding two pipes together on the ground and coupling and sealing 2 set of 10-pipes on the ditch are presented on the following table:

Pipeline construction example

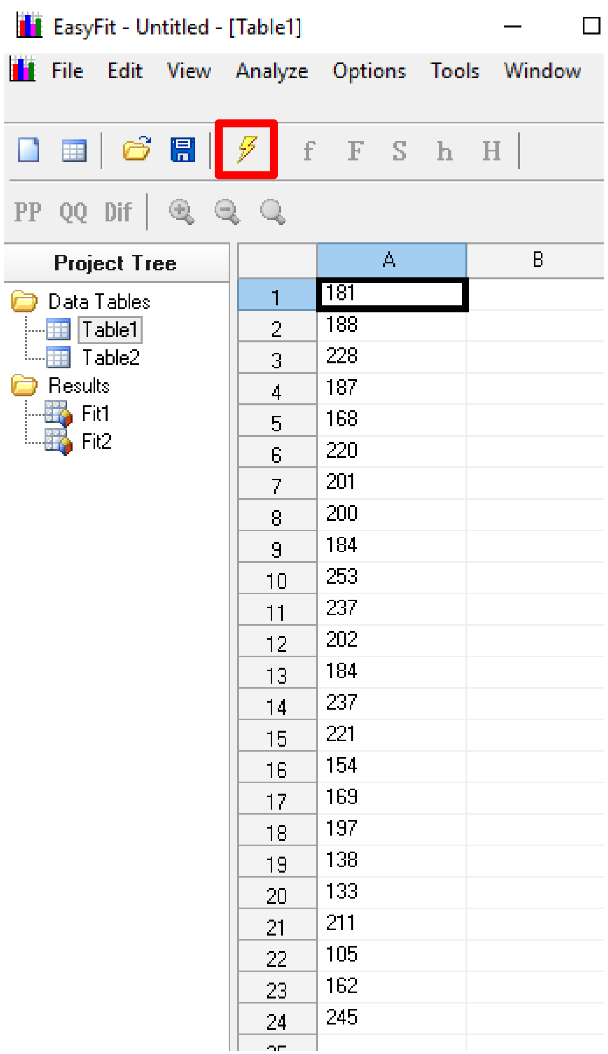
Welding duration for welding two pipes on the ground (minute)
181
188
228
187
168
220
201
200
184
253
237
202
184
237
221
154
169
197
138
133
211
105
162
245

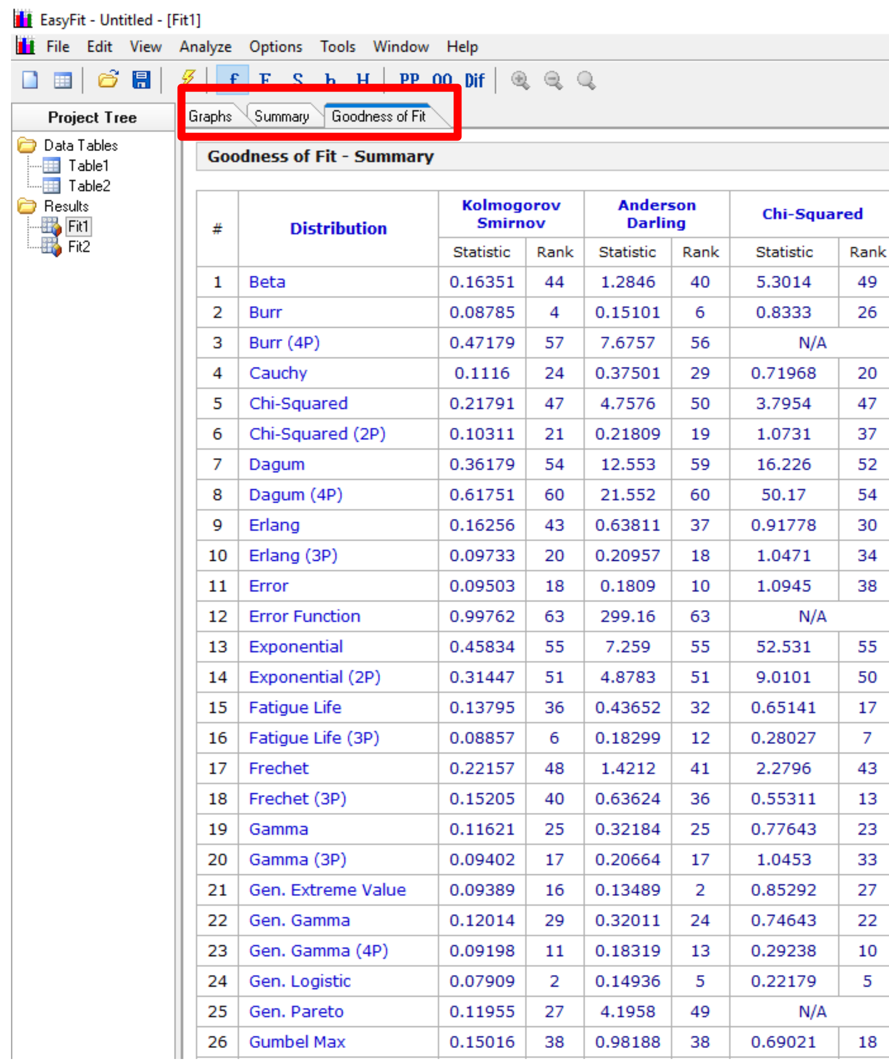
Coupling duration on the ditch (minute)
272
202
238
312
358
196
261
263
346
375

Pipeline construction example



Pipeline construction example





#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Beta	0.16351	44	1.2846	40	5.3014	49
2	Burr	0.08785	4	0.15101	6	0.8333	26
3	Burr (4P)	0.47179	57	7.6757	56	N/A	
4	Cauchy	0.1116	24	0.37501	29	0.71968	20
5	Chi-Squared	0.21791	47	4.7576	50	3.7954	47
6	Chi-Squared (2P)	0.10311	21	0.21809	19	1.0731	37
7	Dagum	0.36179	54	12.553	59	16.226	52
8	Dagum (4P)	0.61751	60	21.552	60	50.17	54
9	Erlang	0.16256	43	0.63811	37	0.91778	30
10	Erlang (3P)	0.09733	20	0.20957	18	1.0471	34
11	Error	0.09503	18	0.1809	10	1.0945	38
12	Error Function	0.99762	63	299.16	63	N/A	
13	Exponential	0.45834	55	7.259	55	52.531	55
14	Exponential (2P)	0.31447	51	4.8783	51	9.0101	50
15	Fatigue Life	0.13795	36	0.43652	32	0.65141	17
16	Fatigue Life (3P)	0.08857	6	0.18299	12	0.28027	7
17	Frechet	0.22157	48	1.4212	41	2.2796	43
18	Frechet (3P)	0.15205	40	0.63624	36	0.55311	13
19	Gamma	0.11621	25	0.32184	25	0.77643	23
20	Gamma (3P)	0.09402	17	0.20664	17	1.0453	33
21	Gen. Extreme Value	0.09389	16	0.13489	2	0.85292	27
22	Gen. Gamma	0.12014	29	0.32011	24	0.74643	22
23	Gen. Gamma (4P)	0.09198	11	0.18319	13	0.29238	10
24	Gen. Logistic	0.07909	2	0.14936	5	0.22179	5
25	Gen. Pareto	0.11955	27	4.1958	49	N/A	
26	Gumbel Max	0.15016	38	0.98188	38	0.69021	18

Pipeline construction example

- Welding Duration: Normal distribution
  - Mean=191.88 minute
  - Standard Deviation=37.205 minute
- Coupling Duration: Uniform distribution
  - a = 172.67 minute
  - B = 391.93 minute

Pipeline construction example

- When in-ditch coupling is completed for a set of pipes, a dozer backfills the dirt to the ditch. The dozer backfills dirt on every kilometer of ditch in one to two days, according to the condition of the route.
- For doing this job we have two options:
  1) Considering whole project in one section with one shovel, one welding crew and one dozer
  2) Dividing the project into two 15 km sections with two shovels, two welding crews and one dozer

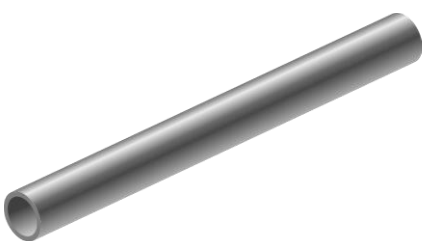
Pipeline construction example

- Main assumptions
  - Every week has 5 working days
  - Working hours starts from 7 am to 5 pm;
  - Every day saved in project duration will save 20 MT of the owner;
  - Renting a shovel costs the project 3 MT/ day;
  - Daily salary of the welding crew is 1 MT
  - Dozer charge is constant and will have no change from one option to the other.
- Which option you offer to the owner and why?

Pipeline construction example

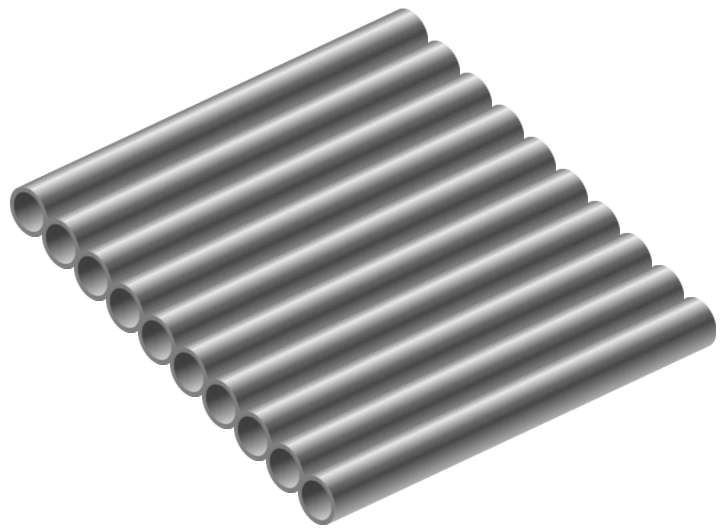
What is Entity in the operation?

One 10m pipe



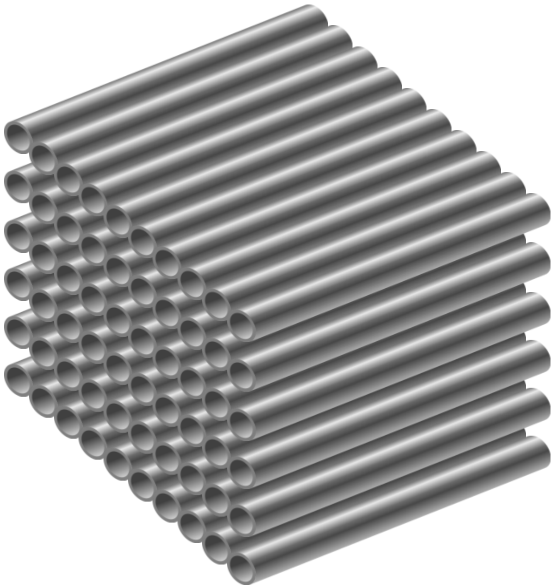
Entity Type 1

Ten 10m pipe



Entity Type 2

30 km of the pipeline



Entity Type 3

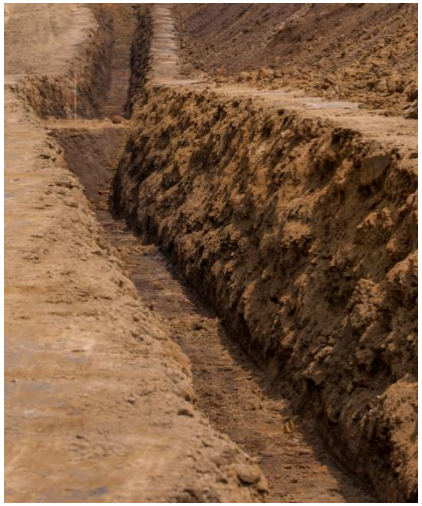
Pipeline construction example

- Any other entities?

One meter of ditch

Entity Type 4

One kilometer of ditch

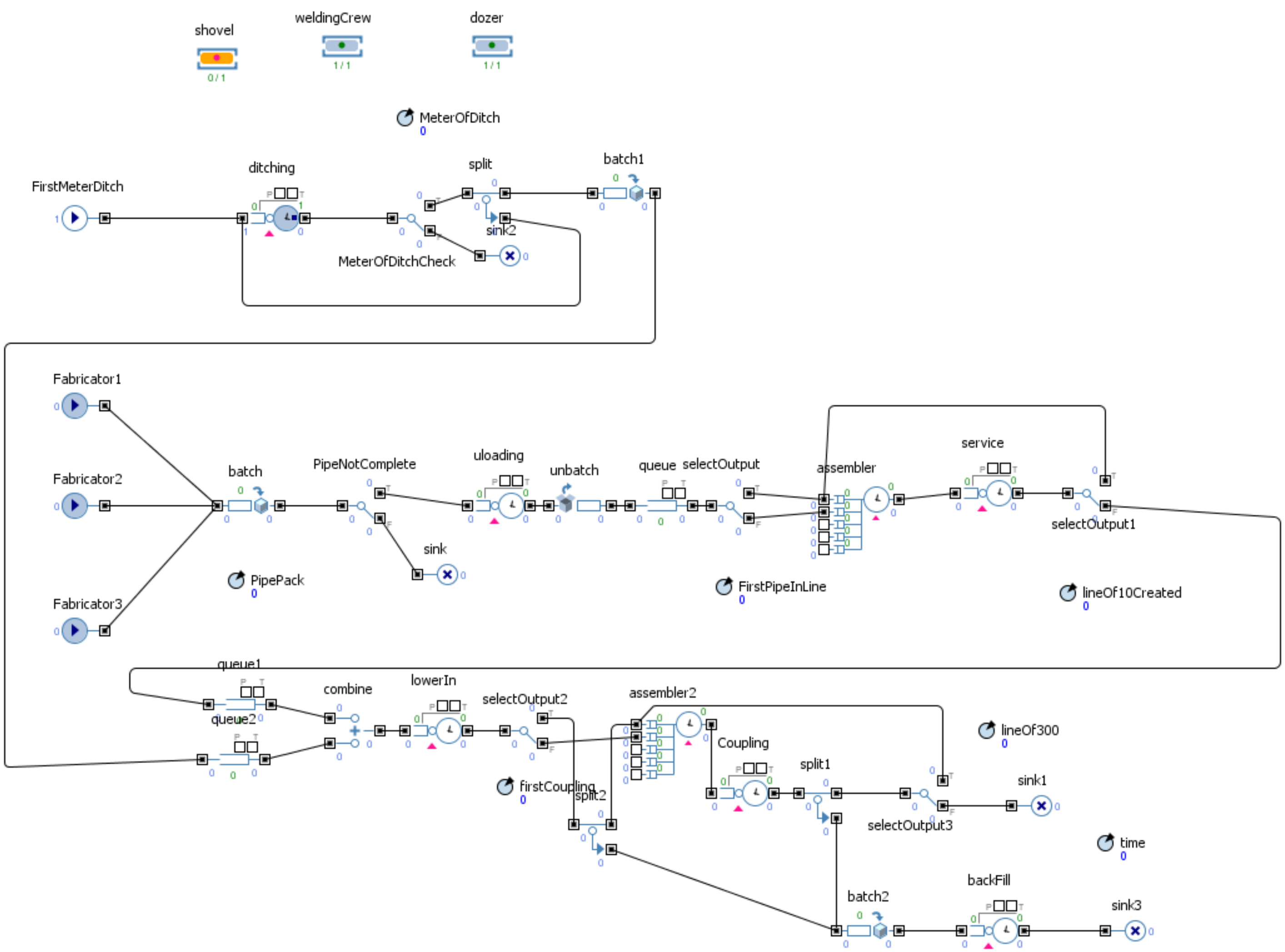


Entity Type 5

Pipeline construction example

- Resources:
  - Shovel, welding crew, dozer
- Activities:
  - Arrival of 3-pipes pack from each fab-shop: Exp (mean=300) minutes
  - Unloading a pack of 3 pipes: Uniform (20,30) minutes
  - Welding pipes: Normal (191.88, 37.205) minutes
  - Digging one meter of a ditches: Uniform(10,15) minutes
  - Lowering 10-pipes into the ditch: Uniform (120,180) minutes
  - Coupling pipes in the ditch: Uniform(172.67, 391.93) minutes
  - Back filling dirt to the 1km of the ditch: Uniform(600,1200) minutes

Pipeline construction example



Home assignment 9

We are going to do a simulation study on a structural steel installation job where 1) structural steel elements arrive at the site from steel fabrication shop 2) structural steel elements are stored in order based on their arrival time by iron worker crew 3) a tower crane moves the elements to their erection location 4) iron worker crew first temporarily stabilizes the steel element on its location to let the crane off and can serve other steel element installation 5) Iron worker crew finishes the installation by bolting and welding the steel elements. There is one crane and two iron worker crews for the operation! The building consists of 500 different steel elements.

Historical input data are presented in the excel sheet attached. You need to develop your own goodness of fit for structural steel element storing and stabilization using Chi square or KS methods based on the number of data provided. Use EasyFit analysis for the rest. Develop a simulation model for the operation using AnyLogic. Explain and submit your data input analysis steps, the developed model, and the achieved results in your assignment report!

Due in two weeks!

Reference

- Banks, J., Carson, J.S., Nelson, B.L. And Nicol D.M. (2004) “Discrete event simulation” Prentice Hall, ISBN: 0131446797., chapter 9.

Thank you!