OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

# Input Modelling

## Statistical Distributions and How to Find Them

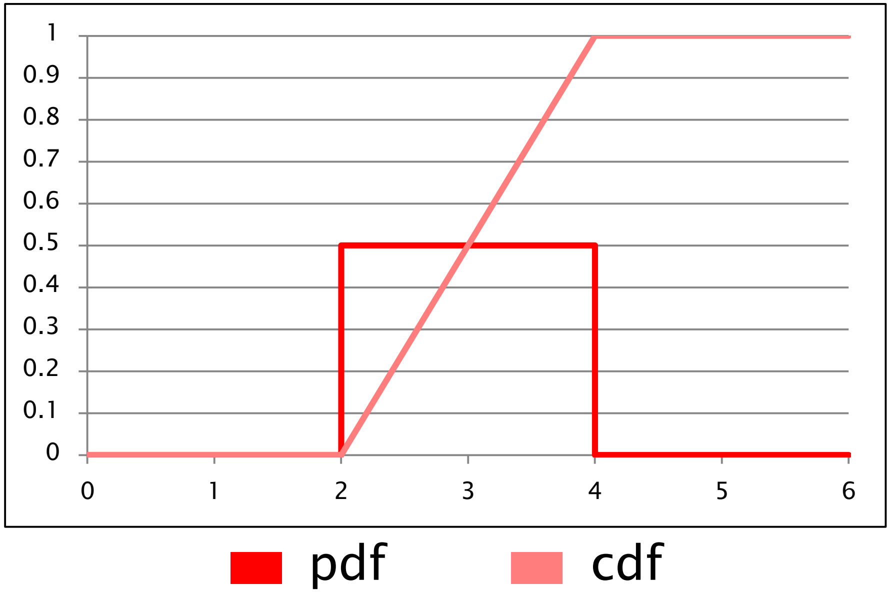
# Review of Common Statistical Distributions

# The Uniform Distribution

## Definition:

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



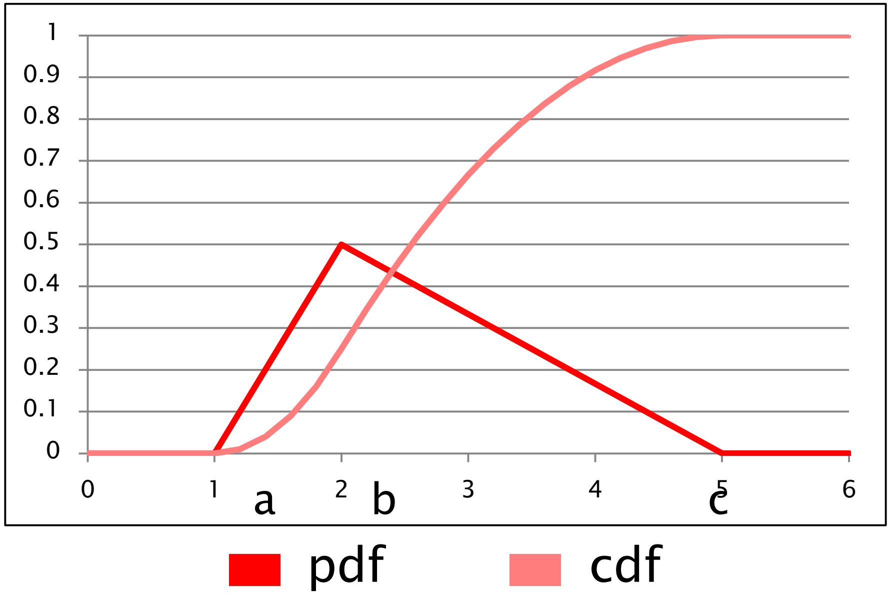

## Applications:

- Random number generation
- When only min. and max. information is available

# The Triangular Distribution

## Definition:

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & a \le x \le b \\ \dfrac{2(c-x)}{(c-b)(c-a)} & b \le x \le c \\ 0 & x \ge c \end{cases}$$




## Application:

- When only min., max. and most common values are known

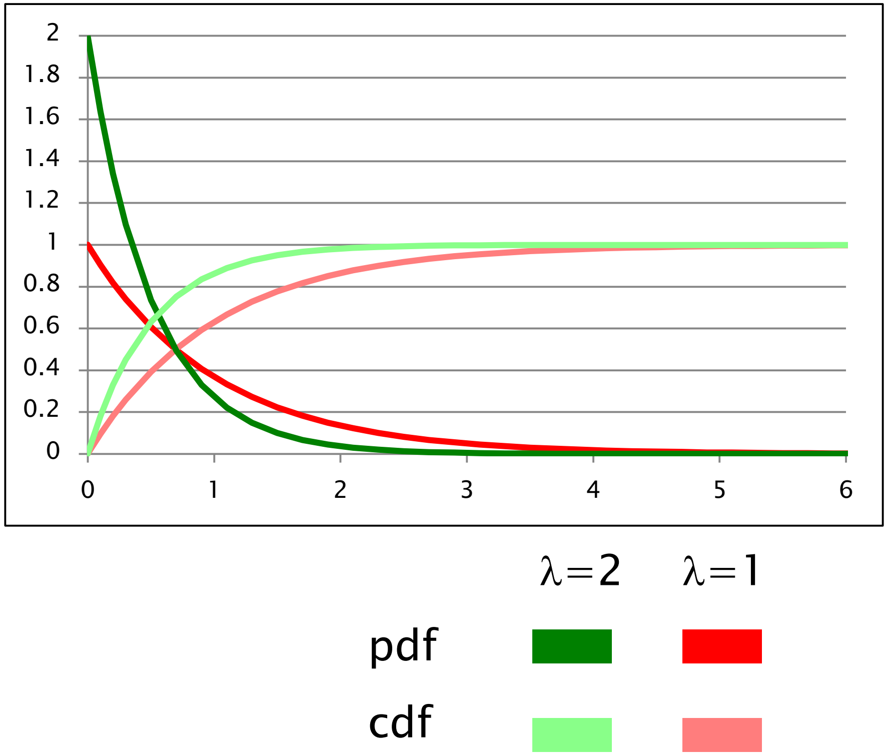
OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# The Exponential Distribution

## Definition:

$$F(x) = 1 - e^{-\lambda x}$$

$$f(x) = \lambda e^{-\lambda x}$$




## Applications:

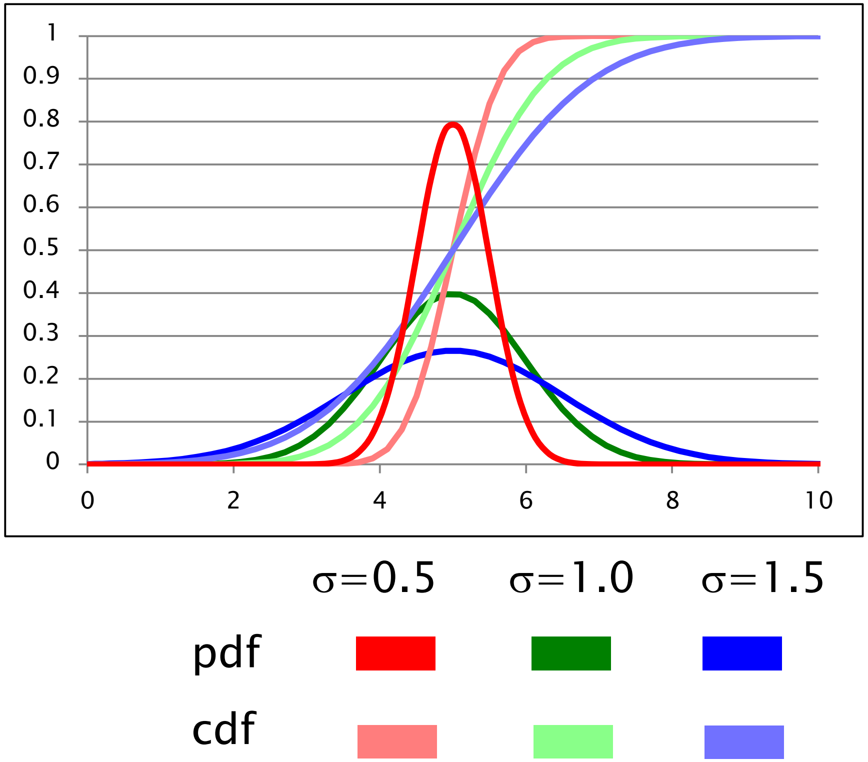
- **Independent arrivals from an infinite population**
- Lifetimes of electronic components

# The Normal Distribution

## Definition:

$$F(x) = ?$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$




## Applications:

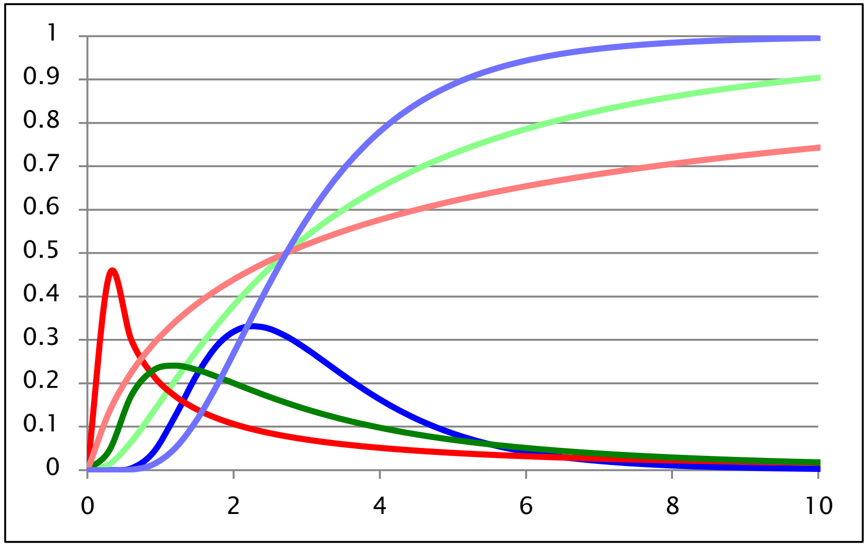
- Shooting at a target
- Common attribute of a population
- Service times

# The Lognormal Distribution

## Lognormal density function:

$$F(x) = ?$$

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$




| | L(1;2) | L(1;1) | L(1;0.5) |
|---|---|---|---|
| pdf | red | dark green | blue |
| cdf | light red | light green | light blue |

## Applications: Many empirically determined distributions

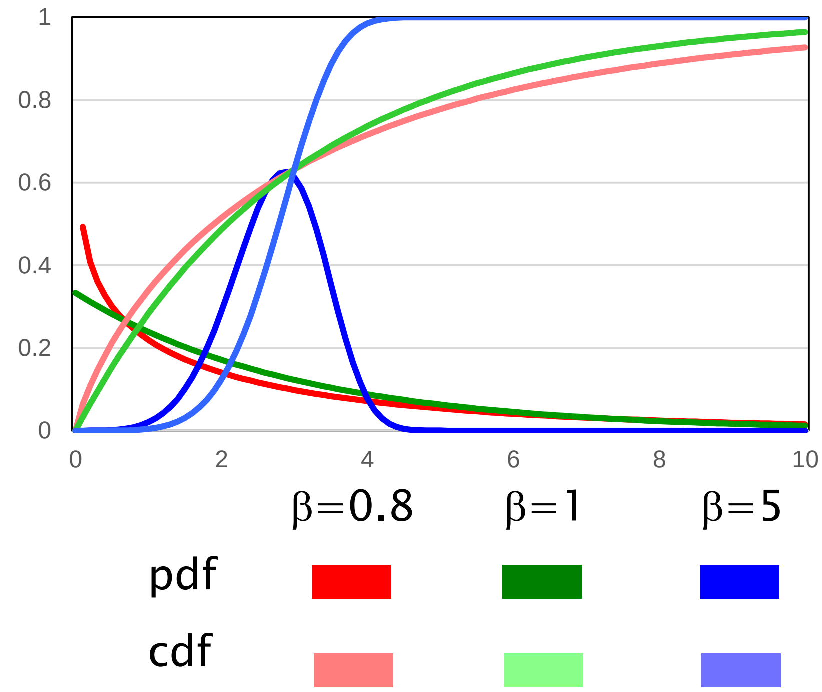
- Household incomes
- Age at first marriage
- Resistance to poison in animals

# The Weibull Distribution

## Definition:

$$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}}$$

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^{\beta}}$$




## Application: lifetimes of different types of components:

- $\beta < 1$: "Infant mortality" (manufacturing errors)
- $\beta = 1$: "Memoryless" (degenerates to exponential dist.)
- $\beta > 1$: "Wear" (mechanical parts)

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# Characteristics of Random Variables

Expected value $E(X)=\mu$:
$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Variance $Var(X)$:
$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx$$

Standard Deviation $\sigma(X)$:
$$\sigma(X) = \sqrt{Var(X)}$$

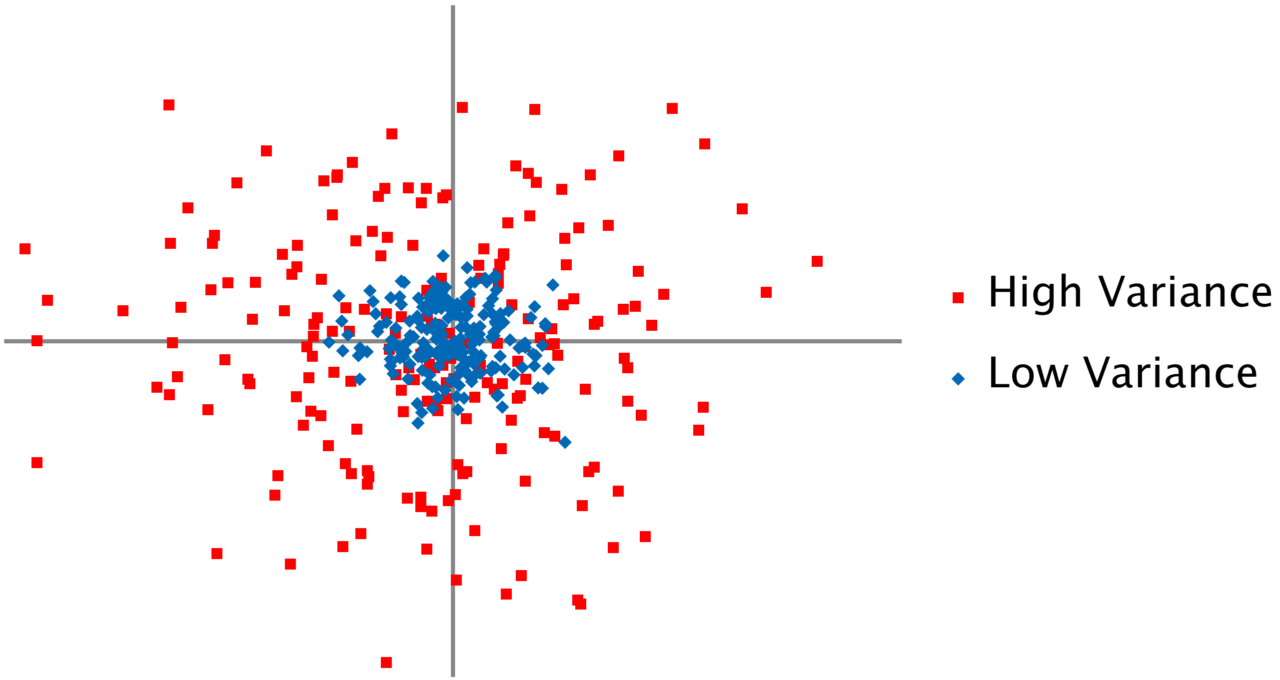
Rule of Thumb for practical distributions:

- $>$75% of all values fall within $\mu \pm \sigma$
- $>$94% of all values fall within $\mu \pm 2\sigma$
- $>$97% of all values fall within $\mu \pm 3\sigma$

# The Effect of Variance

# What is Variance?

A measure of how much values spread, i.e. how far they lie away from their common mean.



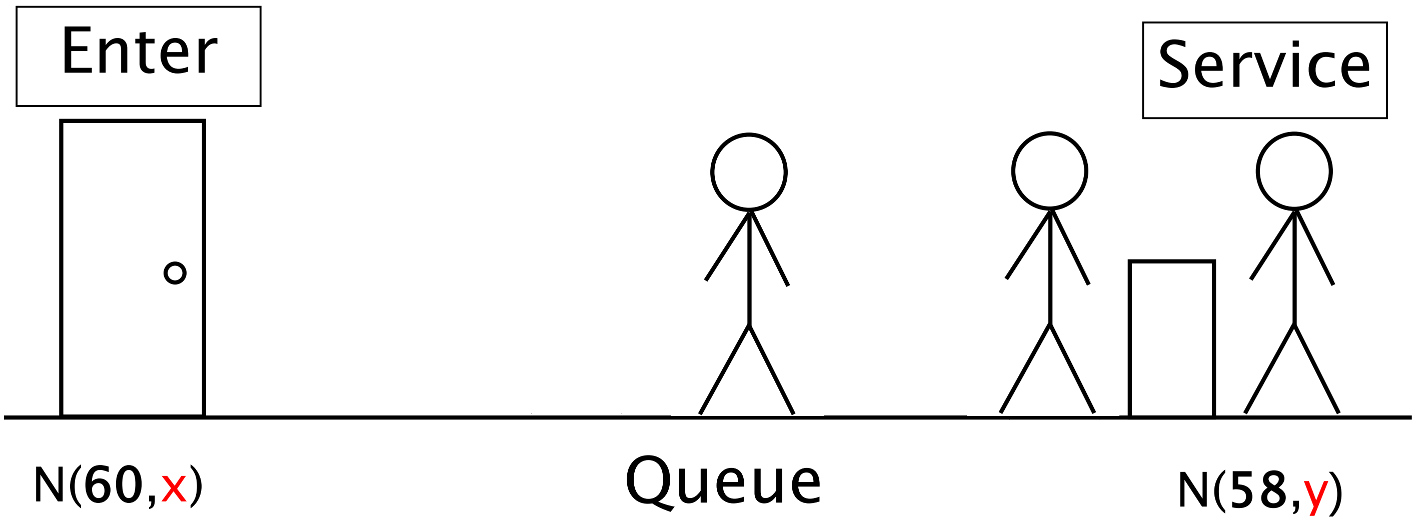

Variance of random variables in a system has a strong influence on the system behavior!

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# The Effect of Variance – Experiments Setup

## A simple queue:

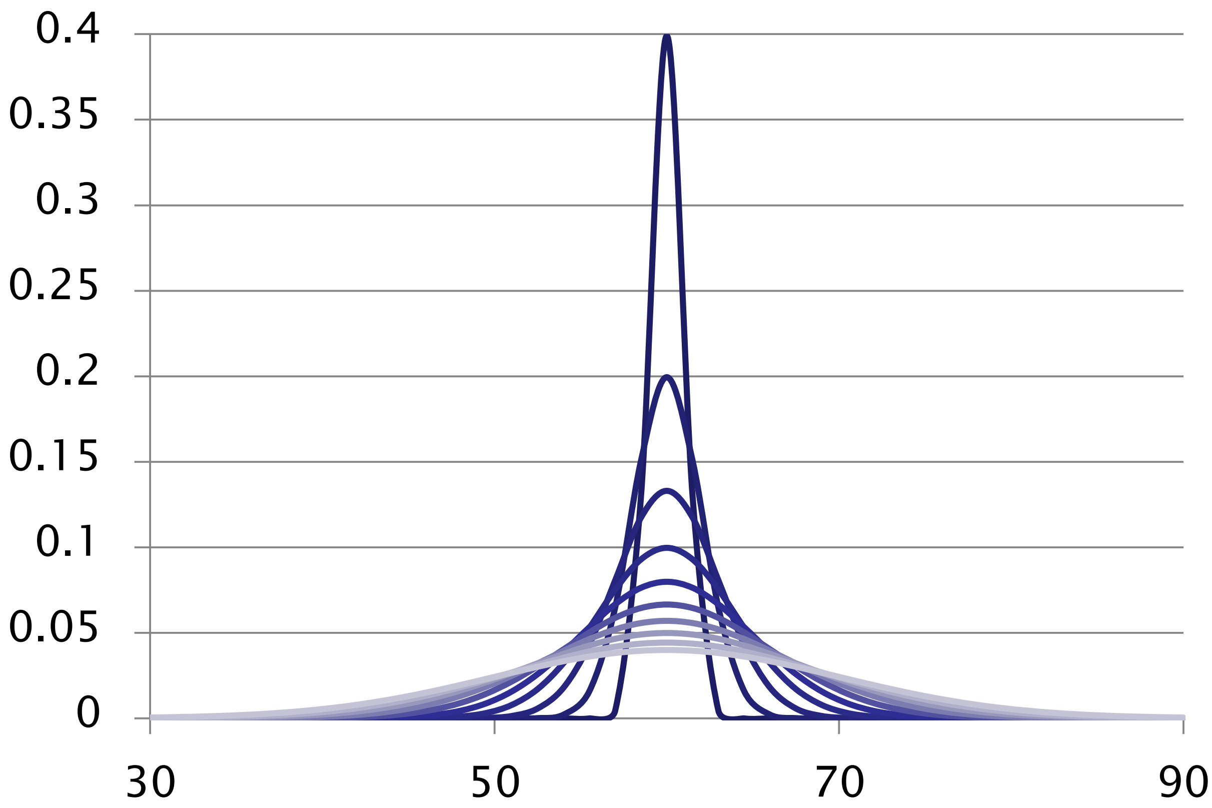
- One server
- One queue with infinite capacity
- Simulation until t=100.000s



OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG | INF FAKULTÄT FÜR INFORMATIK

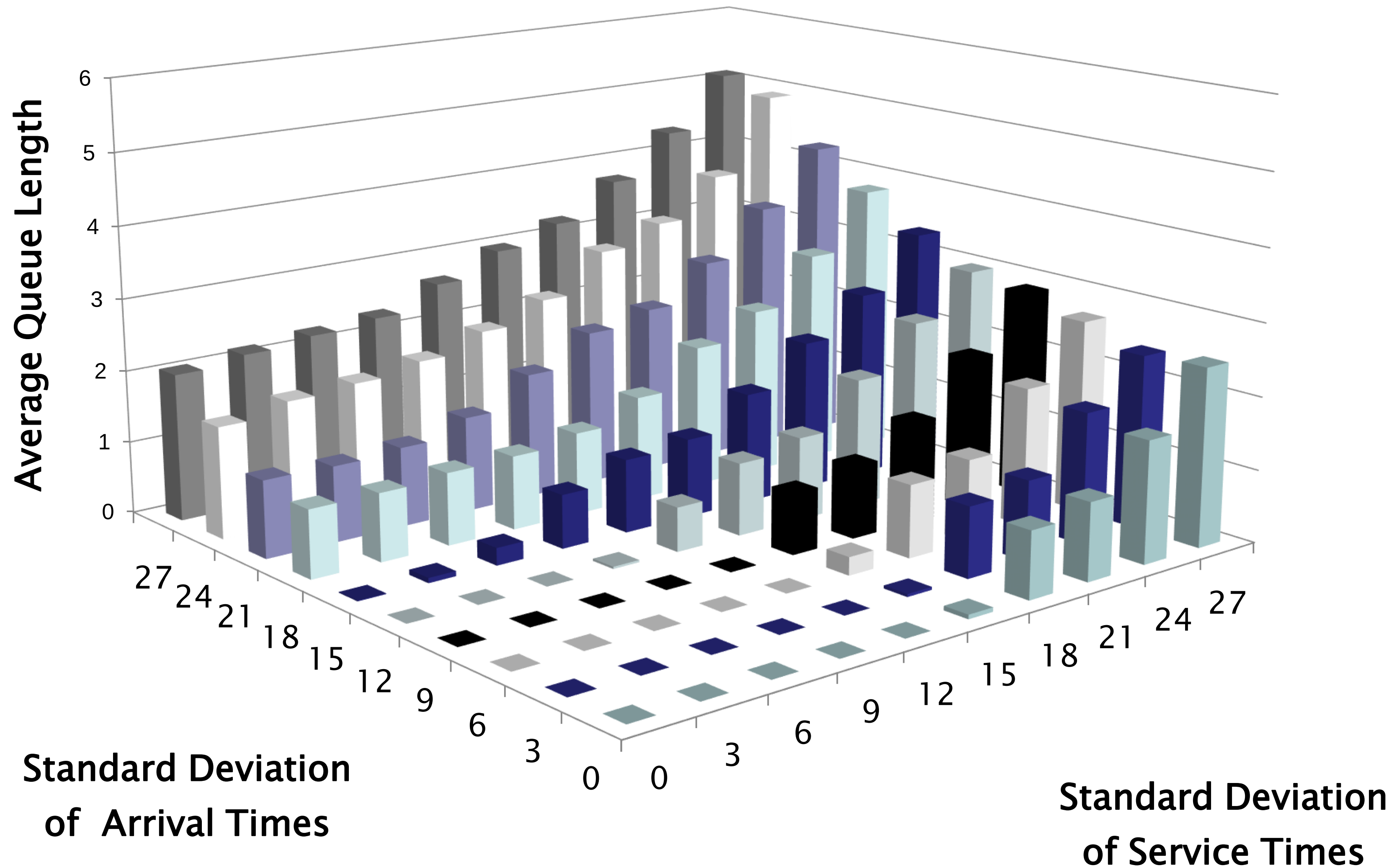
# The Effect of Variance – On Distribution Shape

## Varying standard deviation N(60,1) … N(60,10):



OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

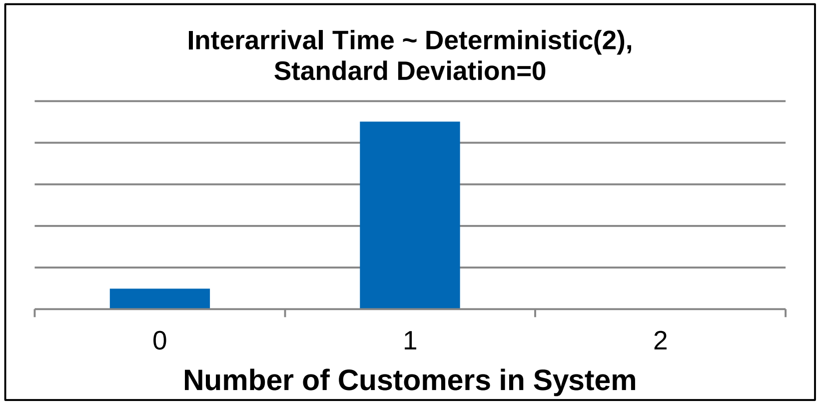
# The Effect of Variance – On *Average* Queue Length

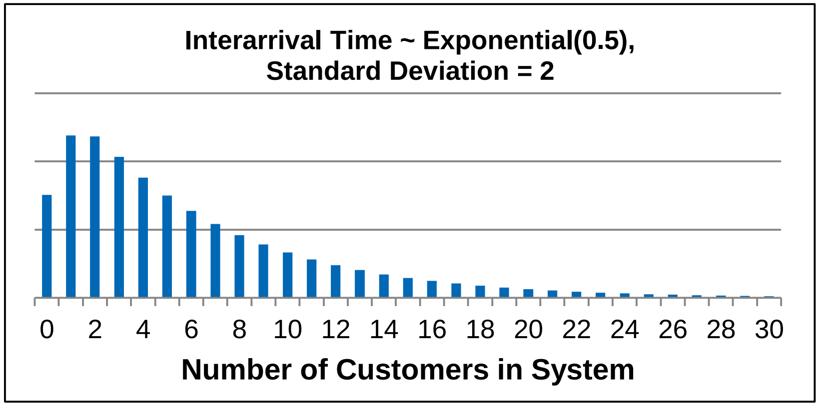


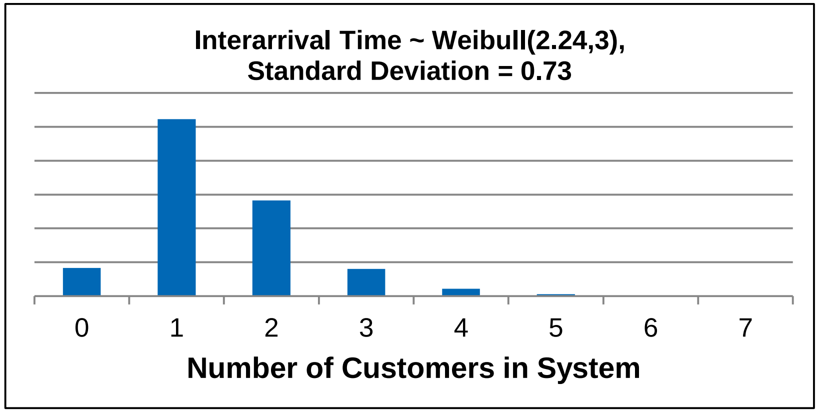
# The Effect of Variance due to Distribution Shapes

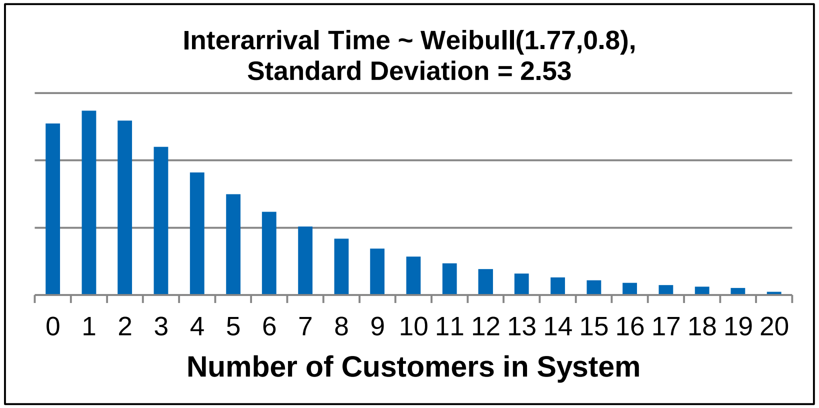
## Service Times ~Deterministic(1.8)

## Average Interarrival Time = 2



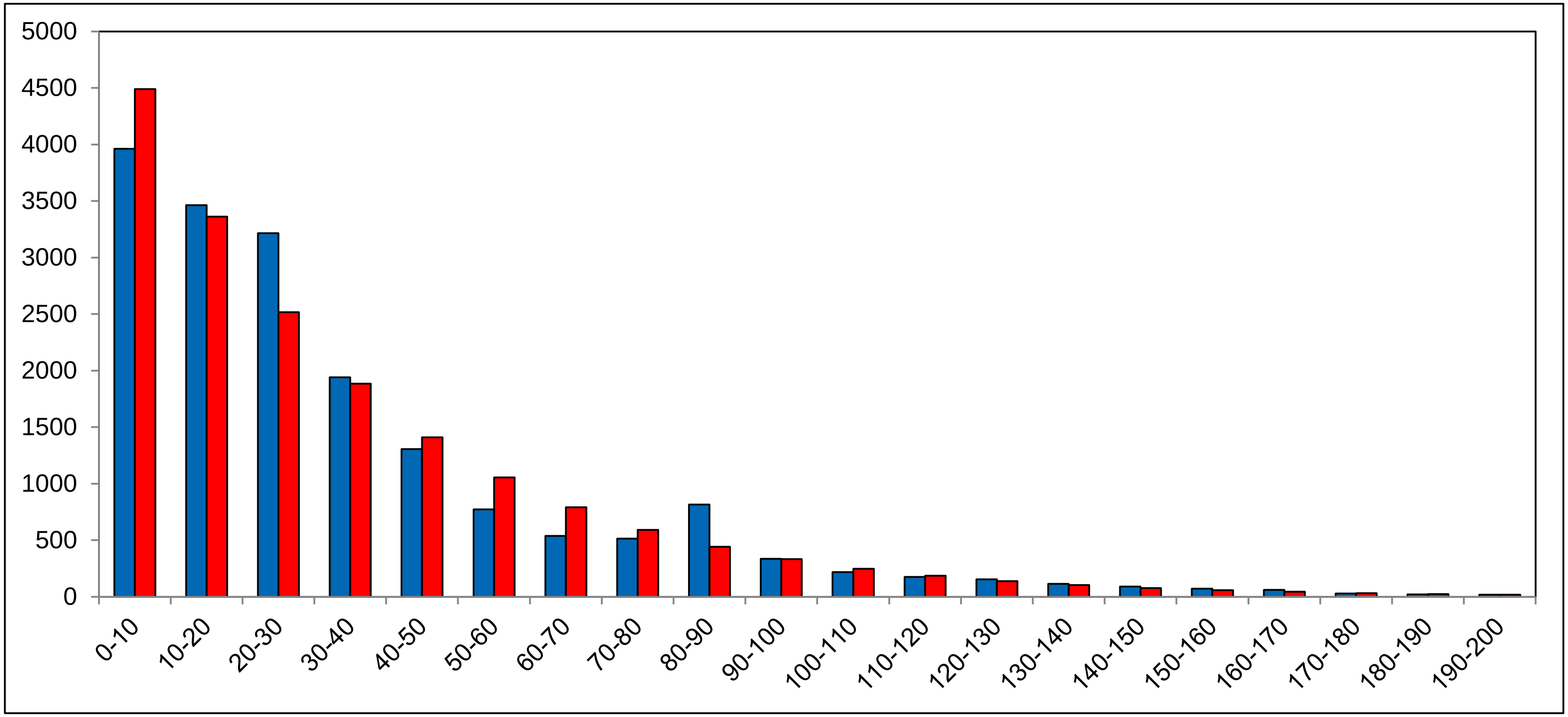



# Applicability of Standard Distributions

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# Example

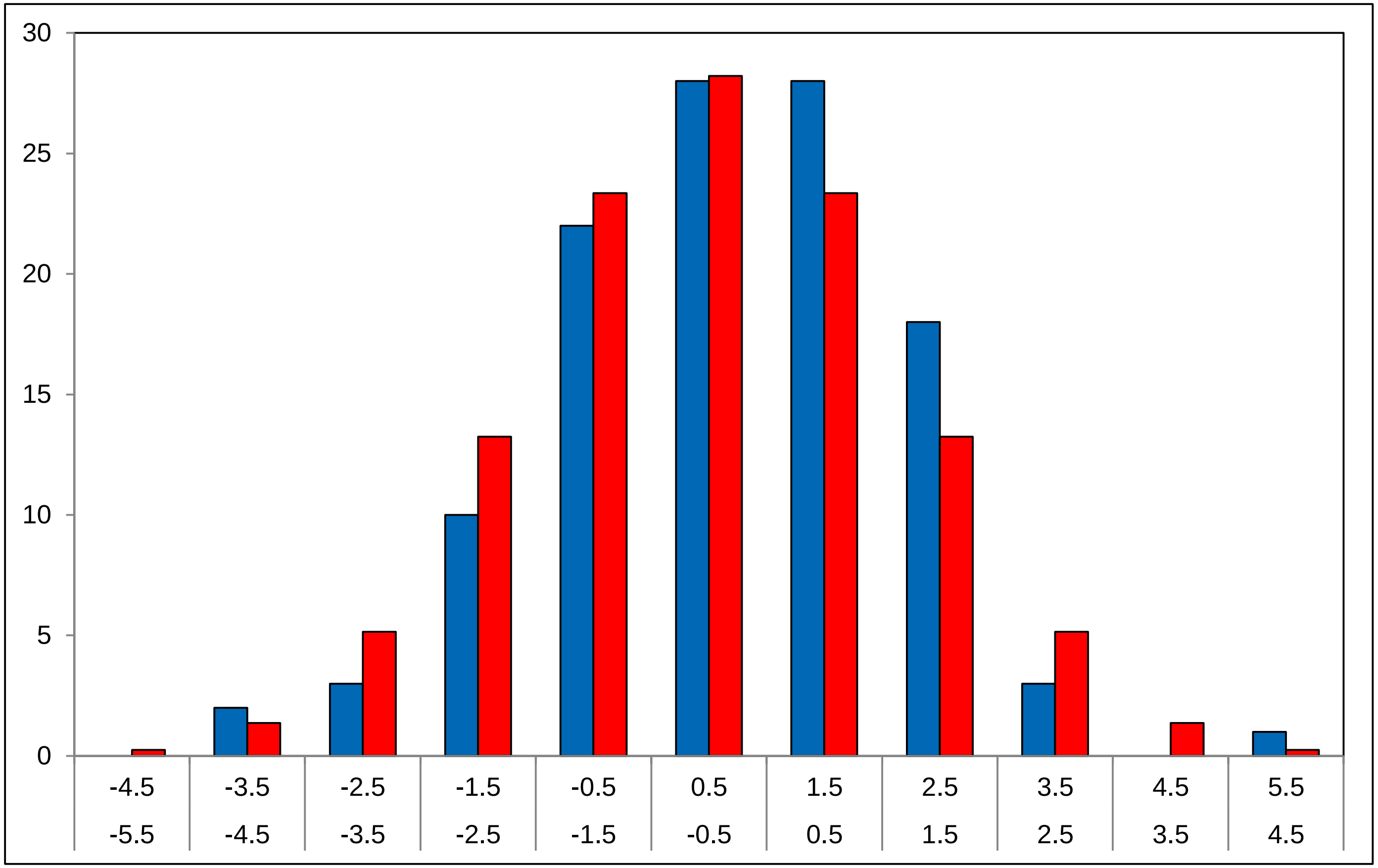
Patients treatment time in days and exp(1/34.55)



OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# Example

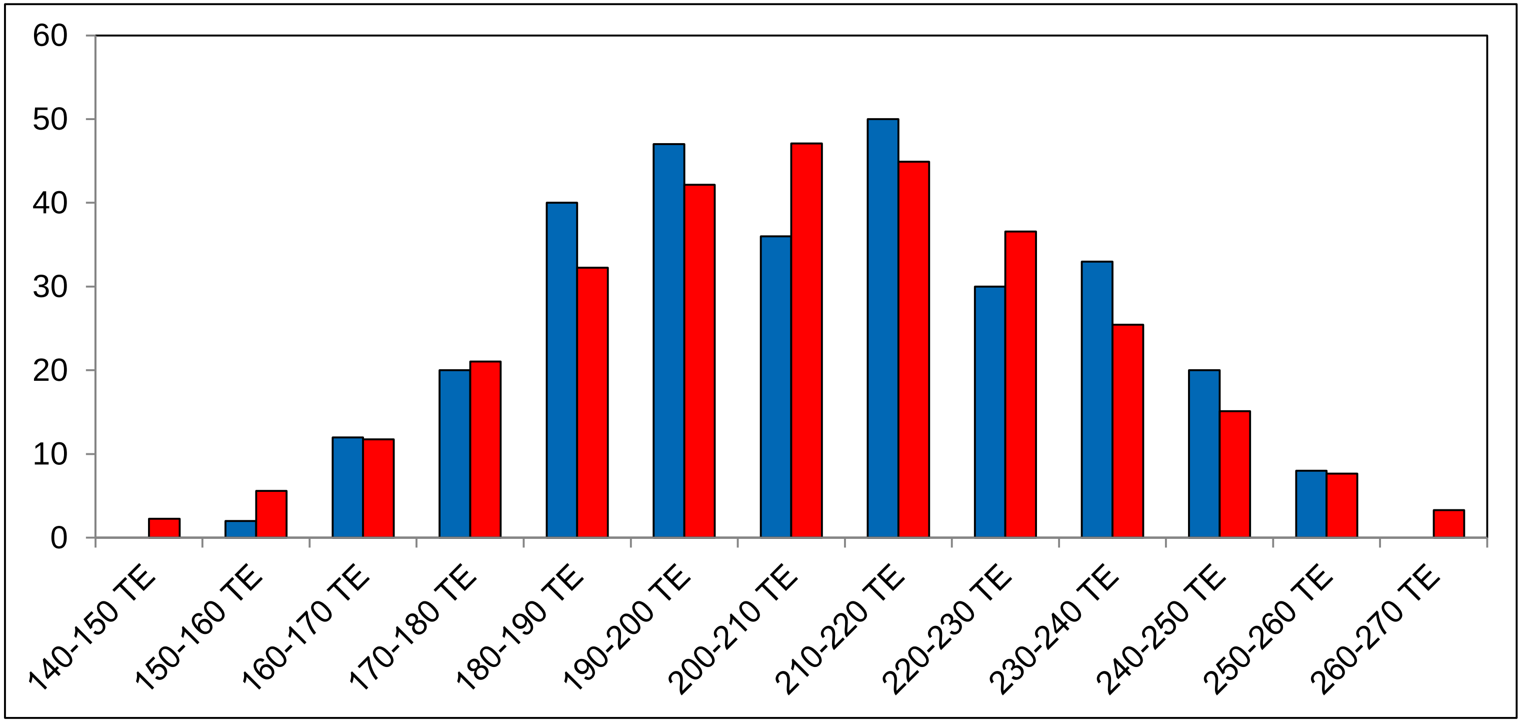
Monthly change of German production index 2000–2009 and N(0,1.6)



OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

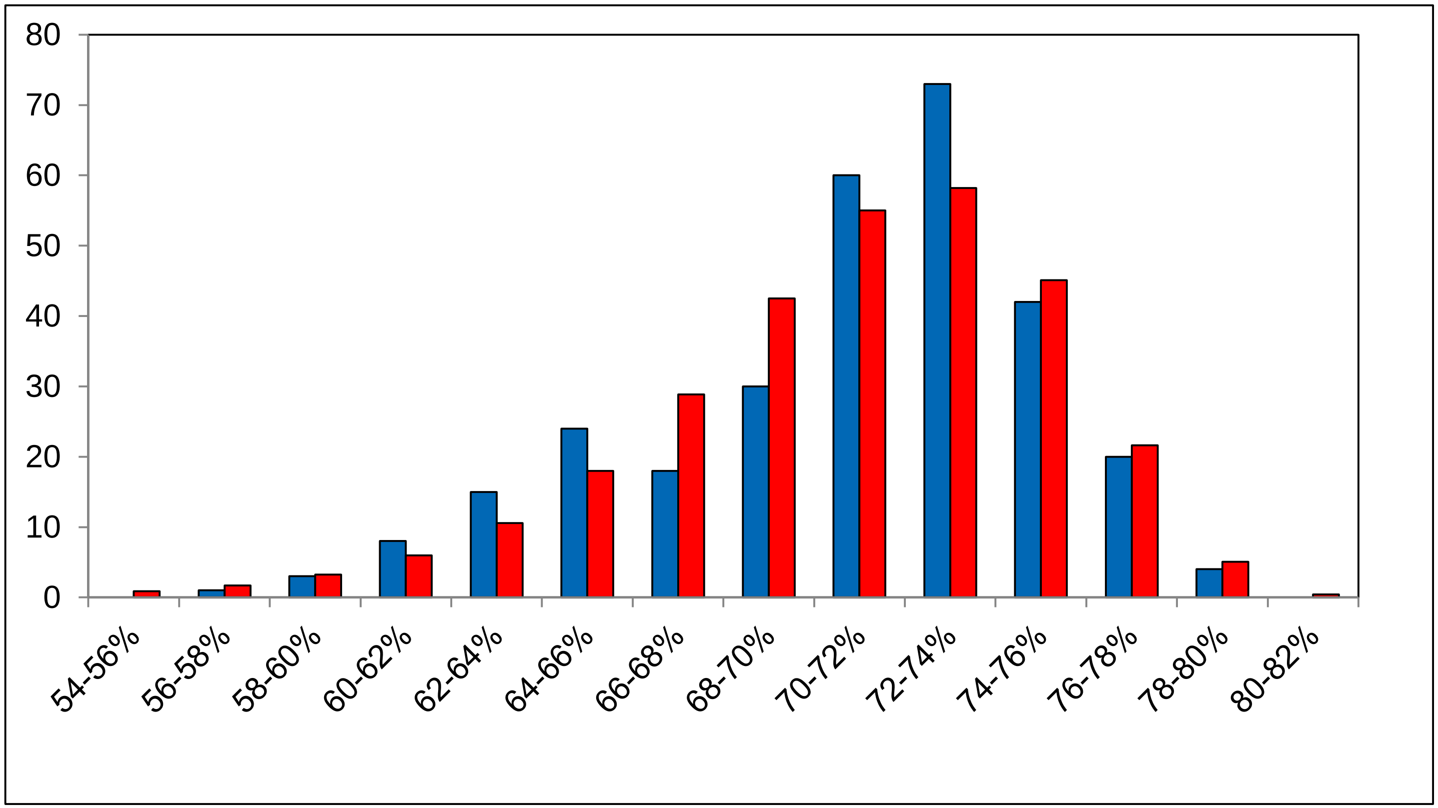
# Examples

## Size of electoral districts for Bundestag election 2009 and N(207 000 , 25 000)



# Examples

Voter turnout in electoral districts for Bundestag election 2009 and W(72.7 , 19.7)



# Fitting a Distribution

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# Input Data

The five steps involved in input data analysis are:

- 1. Collect data (measurement or guesswork)
- 2. Guess a distribution type for the data
- 3. Check the guess for distribution type
- 4. Find the specific parameters for the distribution
- 5. Check the quality of the distribution obtained

# Input Data

## These steps require:

- 1. Use a stopwatch
- 2. Draw a histogram
- 3. Draw a quantile–quantile plot
- 4. Compute parameters from mean and variance
- 5. Do a chi–squared test

# Input Data

## Exercise:

- Find a distribution function describing the age structure of patients suffering from dementia

The computations can be done with a spreadsheet ...

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

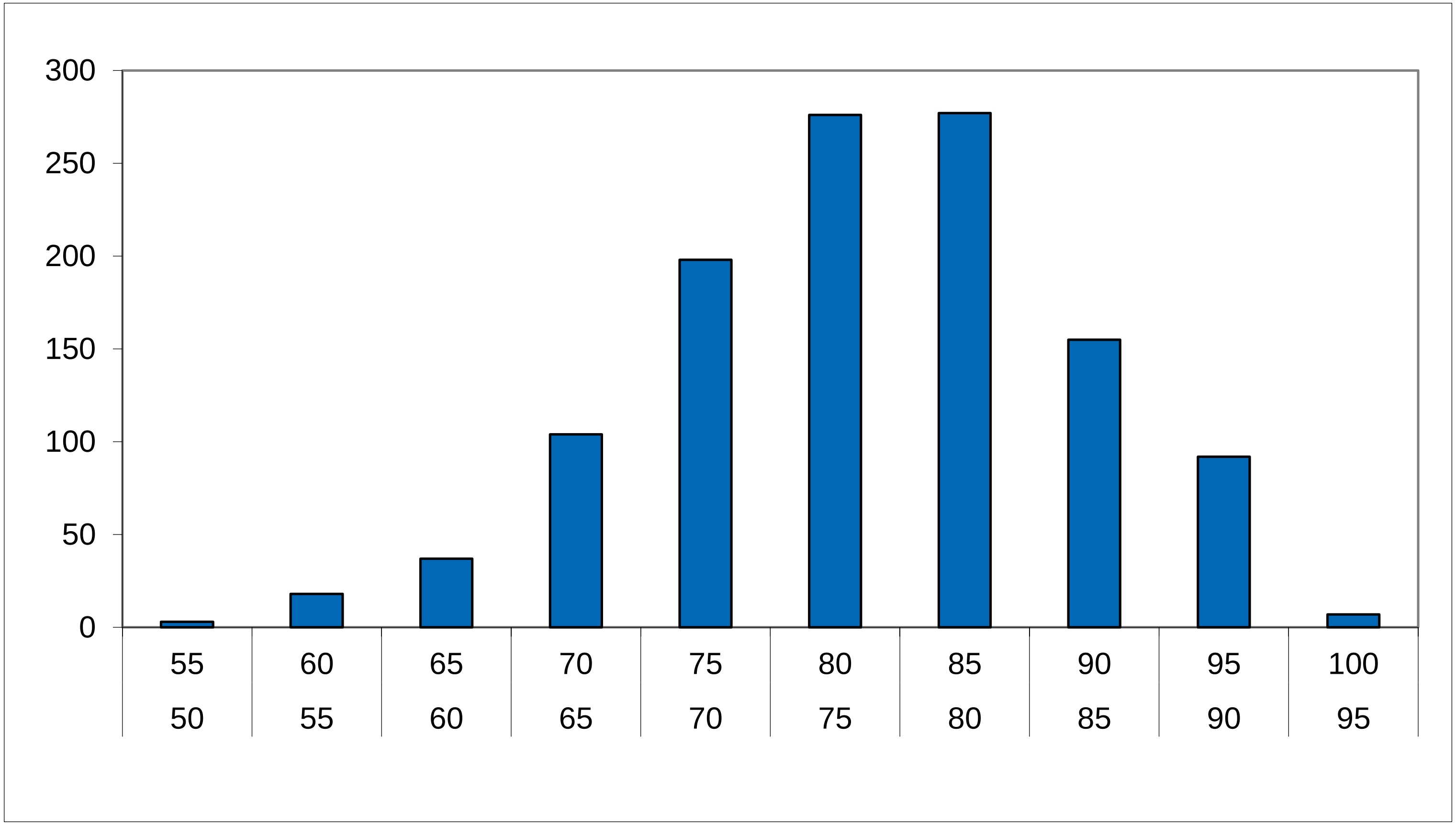
# Input Data

## The measured data:

80 86 68 92 89 73 68 83 74 63 90 83 80 73 85 88 78 74 68 70 71 83 67 86 78 79 86 62 85 88 73 82 80 78 83 78 78 68 75
70 59 85 77 83 72 71 62 81 64 81 89 75 82 77 80 79 82 84 89 65 72 79 80 81 75 74 79 82 81 89 63 73 83 63 82 83 80 75
75 80 75 76 78 74 69 81 76 80 87 71 73 59 94 83 81 71 72 70 79 69 75 68 80 79 90 86 81 77 77 76 87 85 85 82 65 78 70
78 76 86 59 74 79 75 66 88 79 79 88 86 73 85 76 70 77 60 86 74 67 78 76 71 78 77 74 78 72 75 78 81 73 69 76 78 81 83
69 79 77 84 74 81 78 65 69 52 82 76 91 71 81 79 72 69 77 86 70 75 73 78 86 76 72 76 77 74 72 76 76 63 71 75 81 74 54
90 72 85 74 76 72 71 89 78 89 90 69 87 64 82 86 94 83 91 87 70 82 73 78 87 76 75 80 75 69 78 70 63 93 67 74 72 75 67
80 80 76 91 90 73 83 62 83 88 77 79 77 74 81 75 65 79 82 89 80 84 66 70 81 87 72 78 83 76 83 83 70 91 97 76 88 94 88
79 79 73 72 64 75 61 75 78 83 74 55 69 81 82 88 79 71 87 90 82 72 89 69 80 81 67 74 76 67 79 79 73 81 71 85 91 71 75
91 69 76 80 76 67 72 83 90 87 83 68 79 82 74 75 78 76 93 80 87 80 74 89 93 83 82 85 76 75 81 77 75 83 77 91 84 91 75
72 92 82 89 89 83 71 76 70 80 73 81 75 88 69 77 80 68 77 77 83 73 68 81 83 84 61 78 59 72 60 79 72 57 81 77 84 76 78
78 69 76 80 80 65 88 83 75 83 75 81 81 67 69 83 77 80 80 85 79 79 90 84 72 90 87 77 78 75 63 58 78 80 71 78 83 68 87
79 83 75 71 90 92 90 80 85 57 78 81 76 88 57 76 66 72 86 85 77 95 72 70 81 66 77 76 73 90 80 67 84 75 94 67 66 84 76
81 94 75 82 75 86 72 76 79 80 80 76 74 75 87 85 75 80 93 84 84 73 66 84 74 87 67 83 69 84 75 64 90 77 79 92 84 90 82
68 66 84 90 87 72 83 93 65 81 93 92 80 78 80 94 81 75 75 66 54 85 68 80 83 64 90 74 68 78 78 82 84 81 82 78 78 85 73
75 82 76 81 75 86 84 77 77 90 86 81 82 69 84 82 89 81 75 81 84 77 77 71 67 76 79 70 85 77 82 84 92 84 84 77 83 86 82
70 71 71 78 64 78 84 65 93 84 84 89 84 72 82 94 83 73 82 73 70 75 79 89 69 83 83 73 69 86 64 71 73 80 84 90 79 85 75
84 72 80 74 93 73 84 79 77 79 78 57 84 72 87 85 80 81 88 79 92 83 65 73 77 70 87 75 84 91 78 75 85 79 84 70 77 77 84
64 85 73 85 69 85 81 74 86 96 58 82 84 70 89 63 90 72 74 70 68 65 75 66 71 83 85 74 90 79 88 83 78 96 80 79 75 83 89
72 89 80 87 83 81 82 75 71 77 82 71 80 74 80 67 69 90 69 82 80 74 87 91 76 73 83 79 76 81 73 84 76 84 79 77 74 76 71
74 56 77 81 69 66 79 93 72 83 67 78 82 73 74 78 83 86 71 80 74 90 81 67 81 82 82 72 71 70 73 77 69 86 93 71 82 89 80
93 88 92 75 91 82 71 77 75 65 83 82 76 81 60 70 82 81 75 82 80 83 84 88 85 62 85 74 92 68 72 67 81 73 87 72 77 71 62
93 74 93 78 78 72 69 66 94 82 90 86 92 93 83 79 78 66 77 70 63 72 91 77 82 84 78 77 77 76 68 94 59 77 81 83 85 73 78
93 83 80 79 81 66 66 85 76 75 64 70 69 57 93 69 85 83 83 85 88 80 79 86 91 70 80 82 80 76 65 64 79 79 76 87 72 81 76
83 82 80 85 77 86 70 84 86 86 84 81 66 83 88 79 72 71 78 70 79 66 65 76 92 80 86 92 86 70 88 72 78 64 73 69 97 81 80
80 64 69 92 93 83 76 81 84 82 84 76 80 77 87 77 76 92 77 75 66 74 82 85 78 70 94 75 71 74 91 77 79 69 80 83 86 57 90
83 79 64 86 88 76 66 83 92 70 66 64 84 70 85 76 76 73 72 74 82 83 74 92 75 78 86 84 87 79 87 86 73 82 64 84 79 87 68
59 77 81 84 83 89 78 74 91 78 68 79 87 76 70 73 56 93 86 57 78 80 71 67 72 74 78 82 96 64 74 84 67 78 80 78 82 70 75
75 77 79 68 74 84 86 85 85 90 81 78 72 73 72 85 78 82 66 72 85 87 77 84 86 71 73 84 70 85 92 76 89 94 70 69 80 73 84
72 70 78 84 86 86 88 88 61 76 80 80 89 92 61 88 84 74 77 71 77 76 72 77 92 82 93 87 86 77 84 82 89 74 86 70 84 92 74
79 63 70 75 72 92 69 67 68 85 86 83 85 79 87 66 71 91 73 76 74 81 69 89 86 71 74 65 88 80 86 77 97 82 88 87

# Input Data

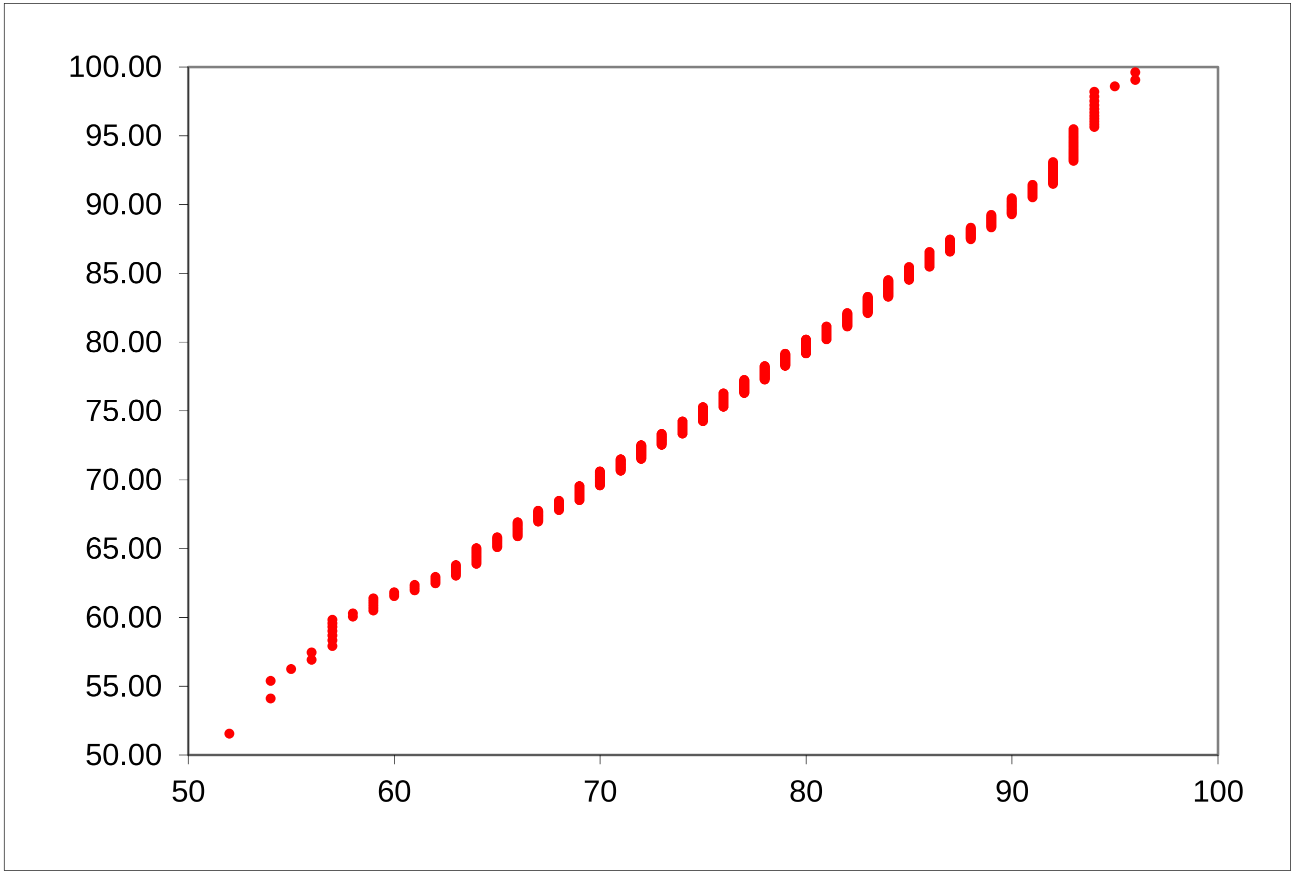
## The histogram:



OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# Input Data

## The Q–Q plot:



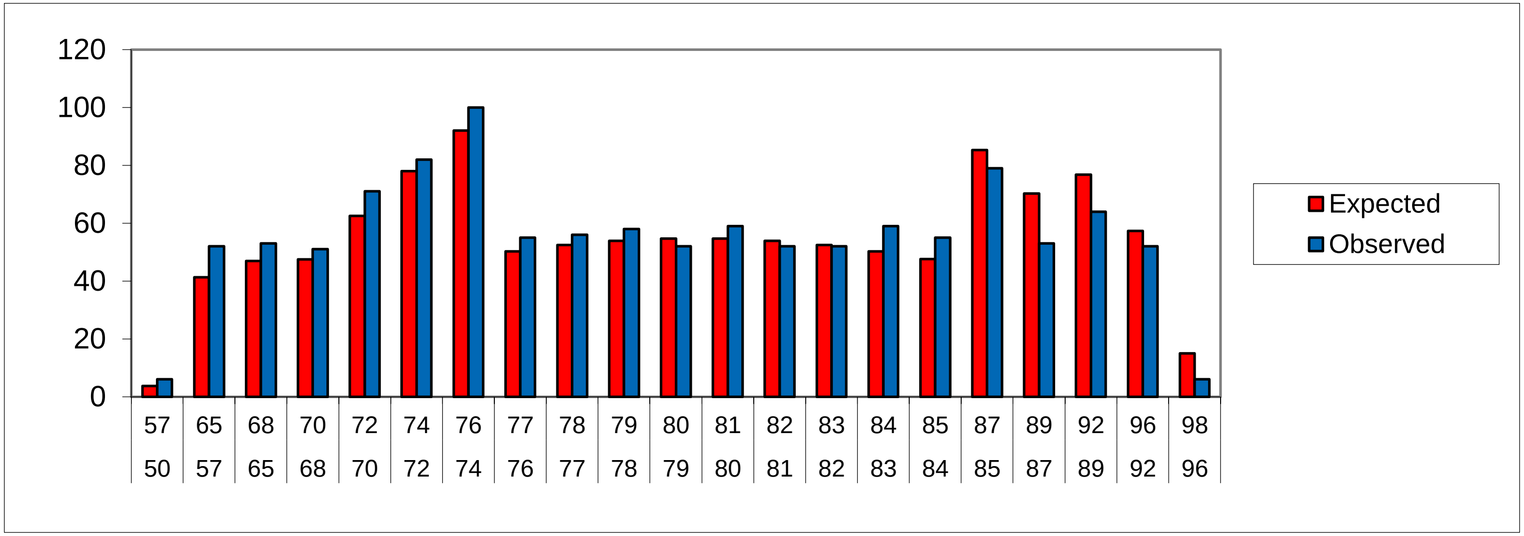
OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

# Input Data

## The Chi-squared test:

| | | xmin | xmax | Expected | Observed | (Ei-Oi)^2/Ei |
|---|---|---|---|---|---|---|
| | | 50 | 57 | 4 | 6 | 1.38 |
| f | 18 | 57 | 65 | 41 | 52 | 2.76 |
| alpha | 0.05 | 65 | 68 | 47 | 53 | 0.79 |
| chisq | 28.869 | 68 | 70 | 47 | 51 | 0.26 |
| Result | ACCEPT | 70 | 72 | 63 | 71 | 1.14 |

| | | | | |
|---|---|---|---|---|
| 96 | 98 | 15 | 6 | 5.40 |
| | | | | |
| | | 1147 | 1167 | 24.20 |



# Learning Goals

## Unknown

- times for treating crew members
- employments duration of IT–consultants

## Given

- List of samples from both

## Draw the Q–Q plot for guessing a distribution type and do the Chi–squared test for the check