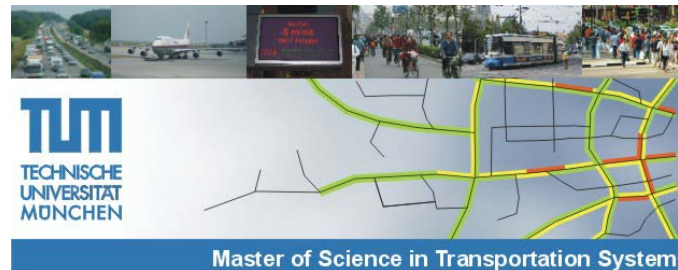


## M.Sc. in 'Transportation Systems'



# Applied Statistics in Transport

Prof. Regine Gerike

Technische Universität München, mobil.TUM

[regine.gerike@tum.de](mailto:regine.gerike@tum.de)

Munich, 24/01/2012

# Plan for Today's Lecture: Repetition

---



- Main Topics
- Exam Last Year

## Organisation: Exam

- After 1 semester
- 17 February 2012, 10:00 – 12:00, room tbc
- Written test
- 60 minutes
- Allowed material: Pocket Calculator, one hand-written A4-page (both sides)

- **Probability:** Classical definition  $P(A) = m/n$  (favourable/possible); How to determine  $n$ , conditional probability, multiplication rule, mutually exclusive – independent
- **Descriptive:** central tendency, dispersion, (position, shape)
- **Distribution:** discrete/continuous
- **Binomial** Distribution: parameters
- **Hypergeometric:** difference to binomial, parameters
- **Poisson** Distribution: parameters, when to apply
- **Normal** Distribution: Standardize the normally distributed values, use the tables for the standard normal distribution
  
- Solve the exercises

- **Inferential:** sampling distribution, central limit theorem
- **Confidence Intervals:** decide when to use normal and when to use t-distribution = small sample size: for mean and proportion
- **Hypothesis:** formulate the appropriate pairs, chose  $\alpha$  depending on the consequences of type I and type II error
- **Statistical Tests:** Gauß-Test, t-test, (not) paired, Wilcoxon, Fisher F-test
- **R:** Simple Commands

- How to read data into R: `read.table`, `load`, `save`
- Basic commands for descriptive analysis: `summary`, `mean`, `var`, `str`, `table`, `hist`, `boxplot`, `plot`, ...
- Commands for computing the pdf/cdf for the binomial, normal distribution:
  - `dbinom(x, size, prob, log = FALSE)`
  - `pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)`
  - `qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)`
  - `rbinom(n, size, prob)`
  - `dnorm(x, mean = 0, sd = 1, log = FALSE)`
  - `pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
  - `qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
  - `rnorm(n, mean = 0, sd = 1)`

## Exam Last Year

- (1) (6 points) Please explain the main characteristics of the scales of variables you know.
- (2) (5 points) The following numbers of inhabitants were registered in the region A:

Year	Inhabitants
1	24,500
2	26,210
3	28,780
4	30,500
5	34,420
6	35,520

Compute the middle growth factor and explain which measure of central tendency you use.

## Exam Last Year

- (1) (7 points) A voice communication system for a business contains 16 external lines. At a particular time, the system is observed, and some of the lines are being used. Let the random variable  $X$  denote the number of the voice lines that are in use at a particular time. Assume that  $X$  is a discrete uniform random variable. Compute the expected value and the variance for the random variable  $X$ . Name two main characteristics of the discrete uniform distribution.
- (2) (9 points) Speed measurements at a section of a highway show an average speed  $\bar{x} = 80 \text{ km/h}$  and a standard deviation of  $s = 10 \text{ km/h}$  (parameters for the sample). The empirical frequencies show that the speed is normally distributed, hence, this distribution can be described by a normal distribution with the parameters  $\mu = 80 \text{ km/h}$  and  $\sigma = 10 \text{ km/h}$ .
- a) Which speed  $x_1$  is fallen below with a probability of 0.15?
  - b) Which speed  $x_2$  is exceeded with a probability of 0.15?
  - c) The fastest and the slowest cars (with highest and lowest speed) should not be considered. Which range of speed is covered by the remaining 95% of speed?



## Exam Last Year

- (1) (5 points) The speed limit of 80km/h was introduced on a rural road. Now the average speed should be measured in order to evaluate the effects of the speed limit. The following requirements were determined: 95% confidence interval, relative error  $E_r=2\%$ . A preliminary study ( $n=150$  passenger cars) showed:  $\bar{x} = 72.8km/h$ ;  $s = 13.2km/h$ . What is the minimal sample size for meeting the requirements for accuracy?
- (2) (9 points) Suppose that 75 percent of people use seatbelts regularly. Give the R-command for determining the probability that in 100 randomly chosen cars with the same number of passengers, in 70 or less of the cars the people do use the seatbelt. Give the R-command for the binomial and the normal distribution.

## Exam Last Year

- (1) (10 points) Pencils produced in firm A should have an average length of 17 cm. Assume the length of the pencils to be normally distributed with unknown variance  $\sigma^2$ . A sample of 5 pencils is taken to check whether the pencils have the required average length. The following lengths are measured for the 5 pencils: 19.2 cm, 17.4 cm, 18.5 cm, 16.5 cm, 18.9 cm. Formulate the null and the alternative hypothesis for this test problem. Does the average length of the pencils in the sample significantly differ from the required mean of 17 cm at a 1%-level of significance (two-sided)? Formulate your conclusions in a final sentence.
- (2) (3 points)
  - a) Please explain briefly the difference between the covariance and the coefficient of correlation.
  - b) Give one example for a negative correlation.
- (3) (6 points) Explain briefly how to read data into R and give two examples for R-commands that can be used for exploratory data analysis.