Syllabus for STT 315-204
Summer Semester B of 2015

Instructor: Sergei Chobanyan
Office: C-408 WH
Office hours: M, Th 5:10-6:10 PM and by appointment
Statistics Help: A-102 WH
Telephone: 3550318
E-mail: chobanya@stt.msu.edu
Class Meetings: MWTh 3:00 p.m. – 4:50 p.m. in 316 Bessy Hall
Web address https://stt.msu.edu/academics/classpages

What is STT-315?
STT315 is a first course in probability and statistics primarily for business majors. We will cover topics in data analysis, probability models, random variables, confidence intervals, and tests of hypotheses with business applications.

Pre-requisite: MTH 124 or MTH 132 or MTH 152H or LBS 118.


Material to be covered: It includes Chapters 1-7. We plan also to master and use MINITAB or EXCEL STATISTICS.

Homework: Home-works will be assigned regularly in class. They will not be collected, but we recommend you to work out the problems. The quiz and exam problems will be based on them.

Quizzes: There will be four 30-minute quizzes (40 points each) scheduled for July 13, July 20, August 3, August 13.
There will be no make-ups for quizzes without written documentation, but the lowest quiz score will be dropped. The quizzes are used to practice the material taught in the lectures and help you develop problem-solving skills.

**Exams:** There will be a Midterm Exam, and a Final Exam.
- Midterm Exam (120 pts.): Monday, July 27.
- Final Exam (160 pts.): Thursday, August 20.

**Grading:** Grades will be based on a straight scale generated from points accumulated during the semester. The break down of possible points is shown below.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quizzes</td>
<td>120 pts.</td>
</tr>
<tr>
<td>Midterm Exam</td>
<td>120 pts.</td>
</tr>
<tr>
<td>Final Exam</td>
<td>160 pts.</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>400 pts.</strong></td>
</tr>
</tbody>
</table>

Your final grade is computed as follows (this should be viewed as a *tentative* scale which may be modified over the term).

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage Range</th>
<th>4.0</th>
<th>3.5</th>
<th>3.0</th>
<th>2.5</th>
<th>2.0</th>
<th>1.5</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>90-100 %</td>
<td>85-89 %</td>
<td>79-84%</td>
<td>73-78 %</td>
<td>65-72 %</td>
<td>60-64 %</td>
<td>55-59 %</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Disclaimer:** The instructor reserves the right to make any changes he considers academically advisable. Changes will be announced in the class.

**Important dates**

- Class begins: 7/6/2015
- Open adds end (8:00 pm): 7/8/2015
- Last day to drop with refund (8:00 pm): 7/16/2015
- Last day to drop with no grade reported (8:00 pm): 7/28/2015
- Final Exam: 8/20/2015
- Class ends: 8/20/2015
INTRODUCTION: BASIC STATISTICAL PROBLEMS

PROBLEM 1. In 1960-s and 70-s American space-crafts Pioneers and Mariners have measured the ratio \( r = M/m \), where \( M \) is the mass of the Earth and \( m \) is the mass of the Moon. The measurements produced the following results: 81.3015, 81.3001, 81.3006, 81.3011, 81.2997, 81.3005, 81.3021.

The calculation of the average \( A = [x(1) + \ldots + x(7)]/7 \) gives \( A = 81.3008 \). The problem consists in finding \( a \) such that with confidence 95% the unknown value \( r \) is between \( A - a \) and \( A + a \).
The method of Confidence Interval allows to show that $a = 0.0008$, in other words, the 95%-confidence interval for the unknown true value of $r$ is in the interval

$(81.3000, 81.3016)$. 
A similar (related to business) Problem 1. The total (population) mean $\mu$ of alcohol content (percentage) in a bottle of French wine of a certain brand should be estimated. The sample mean $m$ of 60 bottles was calculated, $m = 9.3$, $s = 0.8$. Construct a 90% confidence interval for $\mu$. 
**Problem 2.** (Aspirin and heart attacks.) In 1987 NY Times reported that taking aspirin reduces the risk of heart attack. This conclusion was made in collaboration of doctors and statisticians. They have chosen 2 groups of patients: Treatment group (11,037 patients) and Control group (11,034 patients). The patients of the Treatment group were taking aspirin according to a certain regimen, while the patients of the Control group were taking a placebo, a neutral liquid having the appearance of aspirin. During subsequent 2 years 104 patients of Treatment group and 189 patients of Control group experienced heart attack. The medical doctors and statisticians have solved 2 major problems: (a) they have chosen 2 equally likely groups of patients, and (b) on the bases of above figures came to the conclusion that taking aspirin indeed reduces the risk of heart attack.
Problem 3. (The German tank problem). This is a historical problem. During World War II, production of German tanks such as the Panther was accurately estimated by Western Allied scientists using statistical methods. Estimates for some specific months is given as:

<table>
<thead>
<tr>
<th>Month</th>
<th>Statistical estimate</th>
<th>Intelligence estimate</th>
<th>German records</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 1940</td>
<td>169</td>
<td>1000</td>
<td>122</td>
</tr>
<tr>
<td>June 1941</td>
<td>244</td>
<td>1550</td>
<td>271</td>
</tr>
<tr>
<td>August 1942</td>
<td>327</td>
<td>1550</td>
<td>342</td>
</tr>
</tbody>
</table>

Using the statistical language, this is the problem of point estimation of a parameter. The figures in the first column were found by means of the formula:

\[ \hat{N} = m(1 + \frac{1}{k})^{-1}, \]  

where \( k \) is the size of the sample (of destroyed or captured German tanks produced during the period in question) and \( m \)
is the maximum serial number of the tanks of the sample.
Median of Numerical Data

Example. The magazine Forbes publishes annually a list of world’s 20 wealthiest individuals possessed (in billions of dollars) in no particular order

33  26  24  21  19  20  18  18  52  56
27  22  18  49  22  20  23  32  20  18

Stem-and-leaf diagram for the Forbes data

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88889</td>
</tr>
<tr>
<td>2</td>
<td>0001223467</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
</tr>
</tbody>
</table>
What is the median of the Forbes data?

(22)

18   18   18    18  19   20   20   20   21
22
22    23  24    26   27   32   33   49   52
56

**Median** (notation $m$). To find the median of data, arrange it in increasing order $x_1, \ldots, x_n$. If $n$ is odd, then

$$m = \frac{x_{n+1}}{2}$$

(i.e. the median is the middle value of the arranged data). If $n$ is even,

$$m = \frac{x_{n/2} + x_{n/2+1}}{2}.$$

In other words, it is the arithmetic mean of two middle values.

For the Forbes data, $n=20$, and so,

$$m = \frac{x_{10} + x_{11}}{2} = \frac{1}{2}(22 + 22) = 22.$$

The median can be regarded as measure of central tendency.
Another measure of central tendency is the mean:

\[
\bar{x} = \frac{1}{n} (x_1 + ... + x_n).
\]

Make sure that for the Forbes data \( \bar{x} = 25.8 \).

How do you explain the following proposition? The median is resistant to outliers, the sample mean is not. Give an example.

**The variance** of a data is defined as

\[
s^2 = \frac{1}{n-1} ((x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2).
\]

The positive square root \( s \) of the variance is called the **standard deviation** of the data. The standard deviation and variance
are measures of spread of the data. They show how data is spread out of the mean.

Now consider the data

\[ X = 2, 3, 4 \] and shift it by 100

\[ Y = 102, 103, 104 \]

And observe that

\[ \bar{y} = \bar{x} + 100; \quad s_x = s_y. \]

In particular, the variance (and standard deviation) does not change under shifts of data.

Homework: Text: pp. 63-68, Exercises 2.35-2.40, p.69. Answers:

2.36  a. 8.5   b. 25   c. 0.778   d.. 13.44
2.40  a. 2.5, 3, 3;   b. 3.08, 3, 3
c. 49.6