Tentative Schedule:

1. **Sections:** 2.1-2.8

   Topic:
   Sample space, events, combinatorics, probability, conditional probability, Bayes' Rule, independence.

   Exercises: 2.15, 2.37, 2.43, 2.45, 2.60, 2.63, 2.83, 2.89, 2.93, 2.97, 2.101, 2.103, 2.105

2. **Sections:** 3.1-3.5

   Topic:
   Random variable, discrete and continuous probability distributions, joint probability distributions, conditional distributions and independence.

   Exercises: 3.1, 3.5, 3.7, 3.9, 3.13, 3.17, 3.37, 3.38, 3.42, 3.43, 3.50, 3.58

3. **Sections:** 4.1 - 4.3 (until middle of page 129)

   Topic:
   Mean, variance of a random variables, covariance and correlation between pairs of random variables.

   Exercises: 4.12, 4.22, 4.23, 4.37, 4.43, 4.47, 4.53, 4.64, 4.65

4. **Sections:** 5.1 - 5.3 section 5.4, section 5.6

   Topic:
   Discrete uniform distribution, binomial distribution, hyper-geometric distribution, the Poisson distribution and process.

   Exercises: 5.5, 5.35, 5.58, 5.71

5. **Sections:** 6.1 - 6.5 (until top of page 191), 6.7, in chapter 7 from top of page 221 to top of page 222) (theorem 7.11 and corollary 7.1), section 8.6 - 8.7.

   Topic:
   Continuous uniform distribution, normal distribution, normal approximation of the binomial distribution, χ^2 (chi-squared) distribution, *t*-distribution, *F*-distribution

   Exercises: 6.3, 6.10, 6.13, 6.18, 6.20, 6.22, 6.67, in exercise 6.67 find an interval symmetric around the mean, so that the elongation is in this interval with 95% probability

6. **Sections:** 8.1 - 8.2 and sections 8.4 - 8.7.

   Topic:
   Random sample, sample mean and sample variance, sample distribution of sample mean and sample variance

   Exercises: 8.4, 8.7, 8.13, 8.19, 8.21, 8.26, 8.29, 8.39, 8.42,

7. **Sections:** 9.1 - 9.5 (skip subsection on "One-sided Confidence Bounds") and sections 9.12 and 9.14.

   Topic:
   Estimation, confidence intervals and maximum likelihood estimation.

   Exercises: 9.5, 9.7, 9.13, 9.71, 9.72, 9.81, show that the maximum likelihood estimator found in 9.81 is unbiased

8. **Sections:** 10.1 to 10.7

   Topic:
   Statistical hypothesis and test, hypothesis test of the mean in a single sample

   Exercises: 10.19, 10.23, 10.25, 10.15.

9. **Sections:** 10.13, 9.8, 10.8

   Topic:
   Test of variance in a single sample, test of equal mean and variance in two independent samples.

   Exercises: 10.67, 10.73, 10.30, 10.35, for exercises 10.30 and 10.35 calculate a 95% confidence interval for the difference in means

10. **Sections:** 11.1 – 11.3, 11.4 If time allows

    Topic:
    Linear Regression and Correlation

    Exercises: 11.1, 11.4, 11.9, and 11.12,

11. **Sections:** 18.1 - 18.2 If time allows

    Topic:
    Bayes estimators

MATLAB exercise 1. (you are welcome to use another software package):

- Generate a random sample of size 100 from a standard normal distribution using `x = normrnd(0,1,100,1)` (it generate a 100 by 1 matrix of independent samples from a normal distribution with mean 0 and standard deviation 1)
- Create a histogram using `hist(x)` or `histfit(x)`
- Find the sample mean and standard deviation using `mean(x)` and `std(x)`
- Find the size of `x` using `size(x)`, try also `size(x,1)` and `size(x,2)`
- Find the observation that are smaller than 1 using `x(x<1)`, make a hisogram using `hist(x(x<1))`
- Find the number of observation less than 1 using `sum(x<1)`
- Compare `sum(x<1)/size(x,1)` with `normcdf(1,0,1)` - are they close? What is going on here?
- Remark: The above method is known as (simple) Monte Carlo (named after the casino in Monte Carlo)
- Find 95% confidence interval for mean and variance
- Feel free to experiment with the number of observations, mean and variance of the normal distribution, and the threshold

MATLAB exercise 2.

a. Start be downloading data set `wage1.dat` (save as...)
b. In MATLAB import the data using `data = importdata('wage1.dat')` (you might need to adjust the path, alternatively experiment with the `uiimport`)
c. The variable `data` contains three "subvariables":

    i. `data` - a 526 x 24 matrix containing the observed values for 24 different variables for 526 different people.
    ii. `Sectionsdata` - contains the names of the 24 variables
    iii. `colheaders` - same as `Sectionsdata`

    Notice that the first coloumn is named `wage` which contain the wage for the 526 people.

d. For conveinece extract the wage data using `wage = data.data(:,1);`

e. Make a histogram of the `wage` and `log(wage)` using `hist` or `histfit` - which histogram looks most like a normal dsitribution? Why is this relvant?
f. Define `lwage = log(wage);`. We want to test if the mean log-wage is 1.6 at the 5% significance level. Specifiy the relevant statistical hypotheses.
g. Use `mean`, `std` and `size` to calculate the *t*-test statistic
h. Compare the test statistics to the critical values:
i. Assume that `t` is your test statistic. Apply the following commad `2*(1-tcdf(abs(t),size(lwage,1)-1))`. What do you think it calculates?
j. Find a 95% confidence interval for μ
k. Use `ttest` to verify your results

Some of the above in R

l. Data is impoted into R using `data = read.table("wage1.dat",headers=T)`
m. `names(data)` will return a list of variable names. Among these `wage` and `lwage`contain the wage and log-wage
n. Make a histogram of wage and log-wage using `hist(data$wage)` and `hist(data$lwage)`

MATLAB exercise 3.

- We want to test if there is a difference in the (log) wage between men and women
- Use the `wage1.dat` data file again. In MATLAB import the data using `data = importdata('wage1.dat')`
- Extract the 6th column which contains information about gender: `female = data.data(:,6)`
- `female` is a vector of zeros and ones, where one indicates that the person is female and zero indicates male.
- Extract the 22nd column which contains the log-wage: `female = data.data(:,22)`
- Construct `lwage_male = lwage(female==0)` and `lwage_female = lwage(female==1)`. Then `lwage_male` contains the log-wage for males and `lwage_female` contains the log-wage for females.
- Make histograms of the log-wage for men and women, and calculate the sample mean and standard deviation for each group. Does there seem to be a difference?
- Test the following hypotheses, at the 5% significance level:

- H_0: $\mu_1 = \mu_2$
- H_1: $\mu_1 \neq \mu_2$

  You should calculate the test statistic and compare it to ciritcal values and calculate the *p*value.

- Verify your results using `ttest2`.

Bonus info: Analysis of variance (ANOVA)

- Some of you need to compare the means for more than two groups in connection with your projects. For this you need to use ANOVA. This approach is considered in chapter 13.
- Assume there are *k* groups with means $\mu_1,...,\mu_k$. ANOVA tests the following hypothesis:
- H_0: $\mu_1=...=\mu_k$
- H_1: Not all μ_i are equal
- In MATLAB one approach is to have a vector of measurements and a vector indicating which group each measurement is from. Above we have the "measurements" `lwage` and the "group variable" `female`. An analysis of variance can now be carried out using `[h,p,ci,stats] = anovan(lwage,female)`

  How does the results using `anovan` compare to what you obtain using `ttest2`above? Suggestion: Compare *p*-values.