Exercises, lecture 2

12. februar 2010

The exercises today include the last two exercises from last time. Some of the exercises require the use of Matlab. There is a list of useful commands and hints below.

Exercise 1 Let c be a constant. Show that $\text{Var}(cX) = c^2\text{Var}(X)$.

Exercise 2 Let $X_1, X_2, \ldots, X_n$ be independent variables with mean $E(X_i) = \mu$ and variance $\text{Var}(X_i) = \sigma^2$. Show that the average $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$ has mean $E(\bar{X}) = \mu$ and variance $\text{Var}(\bar{X}) = \sigma^2/n$.

Exercise 3 Problem 22.3.5 in the book. (Use calculator and tables in the book, or use basic functions in Matlab - i.e. do not use the commands `ztest`/`ttest`).

Exercise 4 Problem 22.3.10 in the book. (Use calculator and tables in the book, or use basic functions in Matlab - i.e. do not use the commands `ztest`/`ttest`).

Exercise 5 In Matlab:

1. Generate a sample of size 50 from a normal distribution with mean $\mu = 10$ and variance 25.
2. Test the hypothesis $H_0 : \mu = 0$ with both known and unknown variance.
3. What is the confidence interval in each case?
4. Test the hypothesis $H_0 : \mu = 10$ with both known and unknown variance.
5. What is the confidence interval in each case?

Exercise 6 In Matlab: Calculate the 2.5% and 97.5% quantiles in the t-distribution with n degrees of freedom for $n = 10, 100, 1000, 10000$. What is the limiting value as n gets bigger?

Exercise 7 Generate a standard normal $N(0, 1)$ sample of size 1000 in a variable `x`. Multiply each value by 5 and add 100. Generate a $N(100, 25)$ sample of size 1000 in `y` and compare histograms of the two samples (use the command `hist`).

Exercise 8 Use Matlab to verify some of the entries in Table 22.4 in section 22.3 of the book.

Exercise 9 Problem 22.1.1 in the book.

Matlab commands:

In all the commands below we assume `x` is a data sample of real numbers (One can type in a data sample manually like this: `x = [0.7 12 10.2 5.23]`)

- `1:n` generates a vector from 1 to `n`. I.e. `1:4` $= (1, 2, 3, 4)$.
- `x(1:n)` returns the first `n` elements from `x`.
- `m = mean(x)` is the sample average (saved in a variable `m`).
- `v = var(x)` is the sample variance (saved in a variable `v`).
- `n = length(x)` is the sample size (saved in a variable `n`).
- `y = normrnd(mu,sigma,m,n)` generates a normal sample of size `m`·`n` with mean `mu` and standard deviation `sigma` (variance `sigma`^{2}). I.e. `normrnd(10,2,1,100)` generates a sample of size 100 from the normal distribution with mean 10 and variance 4. With `normrnd(10,2,2,100)` 200 samples are returned in a 2 by 100 matrix.
- `q = norminv(p,mu,sigma)` returns the p quantile of the normal distribution with mean `mu` and standard deviation `sigma`. I.e. `norminv(.975,0,1)` $= 1.96$ and `norminv(.025,0,1)` $= -1.96$.
- `q = tinv(p,df)` returns the p quantile of the t distribution with `df` degrees of freedom.
- `[h,p,ci,stats] = ztest(x,m,sigma,alpha)` tests the null hypothesis of `x` being a sample from a normal distribution with true mean equal to `m` and known standard deviation `sigma` at significance level `alpha`. The returned results are:
  1. `h` is 0 if the null hypothesis is accepted and 1 if it is rejected (the alternative is accepted).
  2. `p` this is the 'p'-value which we will discuss next time. Ignore it for now.
  3. `ci` this is the confidence interval for the mean.
  4. `stats` this is the value of the test statistic z.
- `[h,p,ci,stats] = ttest(x,m,alpha)` is very similar to `ztest(·)` above except the variance is unknown. Here the return variable `stats` contains the value of the test statistic t, the degrees of freedom, and the estimated standard deviation.

Matlab hints:

- If you write part of a command and hit the <tab>-button Matlab will help you complete the command.
- If you end a command line with semi-colon the output will not be printed to the screen.
- If a command can return more than one value these are obtained by writing more variable names on the left side of the equation. E.g. `[a b] = command(x)` will return the first value from the command in `a` and the second value in `b`.