Århus November 26 2013

Applied Statistical Analysis with Missing Data
Exercise 1

Consider the dataset `m0_bw`, which holds information on birth weights for 10,000 Danish children— there are no missing data values in this dataset. Let us first explore the dataset:

Part A

Q1: Find the mean birth weight, its standard deviation and standard error.
Estimate a confidence interval for the mean.

Q2: Repeat **Q1** for each gender.

Q3: Compare the birth weight of the two sexes. Estimate the mean difference. Note: Data are not perfectly normally distributed within groups, but let us ignore this for now. Fill in the left part of the table below.

Part B

We will now replace some of the observed birth weights with missing values. First we randomly set observations to be missing with a probability of 30% by running the commands:

```
. set seed ddMM
. replace bweight = . if runiform() < 0.3
```

Where "`ddMM`" is your birthdate. Remember, the function `-runiform()-` generates a random number between zero and one, and the above command thus sets the birth weight to be missing if this random number is smaller than 0.30 - the chance of which is exactly 30%.

Q7: Fill in the right part of the table below. Compare with part A.

Q5: Compare your findings with those of your neighbour. Is the number of missing values exactly the same? Are the means, the se, etc.?

Q6: Discuss what you have just done with your neighbour:
Are the birth weights Missing Completely At Random?
Are the birth weights Missing At Random?
Are the birth weights Missing Not At Random?

	Part A				Part B			
	All	Men	Women	Men- Women	All	Men	Women	Men- Women
n								
mean								
se								
Ci -low								
Ci high								

Part C

Reload the original dataset `m0_bw` before continuing. Now we will assume that the willingness to participate is associated with smoking habits:

```
. generate ProbPart = 0.9 if !missing(cigs)
. replace  ProbPart = 0.8 if cigs == 1
. replace  ProbPart = 0.7 if cigs == 2
. replace  ProbPart = 0.6 if cigs == 3
. replace  ProbPart = 0.5 if cigs == 4

. replace  bweight = . if runiform() > ProbPart
```

Q7: Fill in the left part of the table below. Compare with part A and B.

Q8: Discuss again what you have just done with your neighbour:
Are the birth weights Missing Completely At Random?
Are the birth weights Missing At Random?
Are the birth weights Missing Not At Random?

Part D

Reload the original dataset `m0_bw` before continuing. Now we will assume that the willingness to participate is low (50%) among women, who gave birth to a small child:

```
. generate ProbPart = 1
. replace  ProbPart = 0.5 if bweight < 2500

. replace  bweight = . if runiform() > ProbPart
```

Q9: Fill in the right part of the table below. Compare with part A , B and C.

Q10: Discuss again with your neighbour:
Are the birth weights Missing Completely At Random?
Are the birth weights Missing At Random?
Are the birth weights Missing Not At Random?

	Part C				Part D			
	All	Men	Women	Men- Women	All	Men	Women	Men- Women
n								
mean								
se								
Ci -low								
Ci high								