Lab Worksheet VI

Statistical inference I

This lab worksheet walks you through the very basic concepts of statistical inference (no need for Stata today, pencil and paper will do). For the purposes of this exercise, it will be very helpful if you already know a bit about the following concepts:

- Population.
- Sample.
- Sample statistic.
- Sampling distribution.
- Central Limit Theorem.
- Normal distribution.

The main learning goals is to understand what inferential statistic is, but we will also look at how we can collect random samples from larger populations as well as how we go about constructing sampling distributions using repeated random samples. *Figure 1* below shows the basic "statistical procedure." As you can see, it has three main components: a **population** or target group of interest, a **sample** or sub-group of the target group, and **probability**, which allows us to say things about the population based on a sample. After collecting a sample, we produce some summaries (i.e., we describe and explore the data). However, this does not really speak of the population but of the observed data in a sample. These observed data might differ from the population, so we rely on probability and inference to draw conclusions. Below *Figure 1* there are some questions for you to answer.

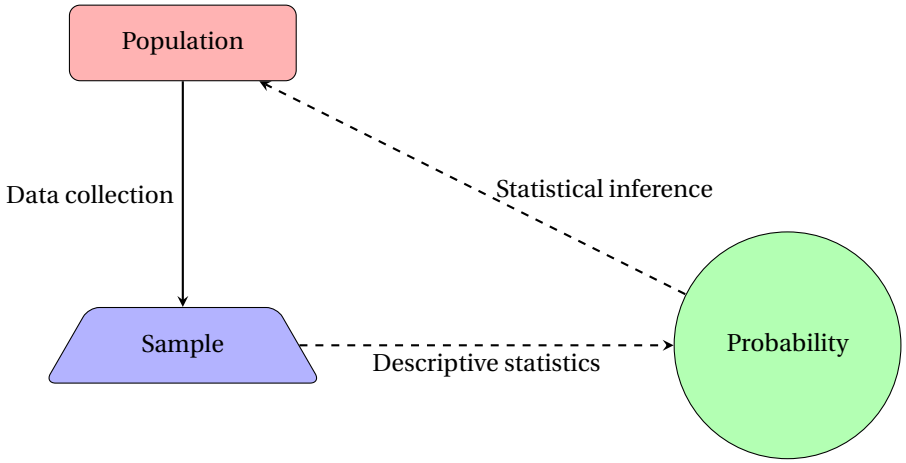


Figure 1: Doing statistics.

? Provide some examples of populations and samples in the box below.

? Might the following samples (simple random sampling) be representative?

Target population:
Lawyers in Greater London area.
Sampling and sample:
All lawyers working in the Greater London area are assigned an X-digit identification number. 500 lawyers are selected at random by using a random-number generator.
How this sample could have been non-representative:
This samples would not be representative if only lawyers with a British passport had been selected (presumably there are lawyers with non-British passports working in the Greater London area).

Target population:
University students studying in Scotland.
Sampling and sample:
Using Scottish universities' student records, a random sample of 1,000 students was pooled. This sample was designed to comprise students older than eighteen and younger or equal than twenty-four. All nationalities, ethnic groups, and sexes were included.
Is this sample representative? Why/why not? How could it be representative?

Target population:
University students studying in Scotland.
Sampling and sample:
50 interviewers were hired in order to interview university students from all Scottish universities. Interviewers were sent to the field only on odd working calendar-days. Data collection was ended after obtaining 1,000 completed questionnaires.
Is this sample representative? Why/why not? How could it be representative?

Target population:
University students studying in Scotland.
Sampling and sample:
Researchers gather a comprehensive list of university students in Scotland. Students are grouped into "male" and"female" categories. Conveniently, 60 percent of university students in Scotland are female. Therefore, the researchers select at random 600 female students and 400 male students.
Is this sample representative? Why/why not? How could it be representative?

Next we are going to construct a sample using simple random sampling. For this, we will use real data from the European Social Survey 2012, but this time we will only focus on respondents from Scotland. On the last page of this worksheet you will find a table with 100 self-reported happiness levels. Happiness was measured from 0 to 10 (being 10 the happiest one can be). Our task is to construct a sample of 15 individuals, so:

? What is the value of N? And the value of n?

? How would you chose a random sample of 15 individuals from *Table 1*?

One "manual" way to construct a random sample involves using a table of random numbers. The numbers in the table below were placed in a random order by a computer, so technically is not purely random (but it is good enough for our exercise). We need to use this table of random numbers to chose 15 individuals from *Table 1* that reports happiness levels in Scotland. First we must devise a "path" or "method", for example: "two to the right, one down, one left, one up, two left, two up, and repeat" (it can be easier, of course). Now we place a finger randomly on the table and simply follow our "path." The figure we end up on will be our first sampled individual. Finally we repeat this procedure as many times as necessary.

? Annotate below the 15 respondents' ID and their happiness levels.

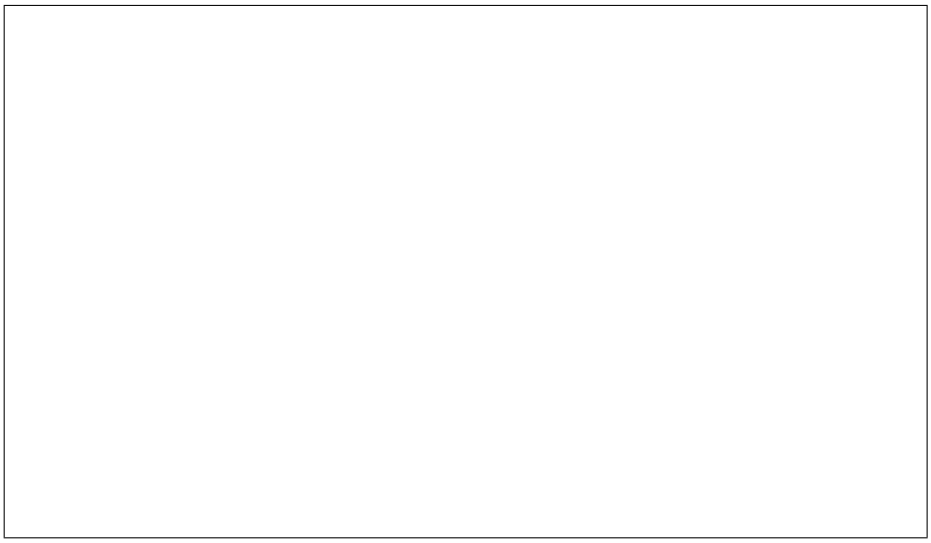
056 034 074 055 037 064 084 051 050 026 026 078 095 019 073 011 071 017 052 075 052
047 035 071 065 009 077 098 009 002 024 057 059 100 003 021 067 030 087 017 079 091
028 044 024 085 075 077 073 083 054 041 054 013 064 036 067 074 080 064 012 068 027
096 088 066 005 087 069 096 016 083 082 057 031 083 026 050 078 042 076 049 057 006
084 079 067 002 096 040 082 030 041 033 056 062 090 032 034 053 072 062 018 048 084
023 060 049 029 090 007 008 005 015 086 072 086 044 069 068 099 006 011 095 043 073
058 028 093 097 037 092 001 027 047 088 014 089 063 015 058 056 010 074 007 080 088
038 089 010 099 034 028 072 014 061 046 038 061 094 022 063 040 058 003 093 050 053
016 055 065 081 060 021 039 040 037 046 091 004 018 076 001 100 031 038 043 027 075
004 090 059 025 029 069 024 032 059 052 019 045 094 094 047 063 087 041 079 039 085
020 070 020 042 030 065 033 077 046 066 078 070 093 025 054 068 071 089 035 025 081
085 048 060 097 012 092 053 044 045 042 051 022 036 049 081 033 031 062 043 048 032
080 036 095 091 056 034 074 055 037 064 084 051 050 026 026 078 095 019 073 011 071
017 052 075 052 047 035 071 065 009 077 098 009 002 024 057 059 100 003 021 067 030
087 017 079 091 028 044 024 085 075 077 073 083 054 041 054 013 064 036 067 074 080
064 012 068 027 096 088 066 005 087 069 096 016 083 082 057 031 083 026 050 078 042
076 049 057 006 084 079 067 002 096 040 082 030 041 033 056 062 090 032 034 053 072
062 018 048 084 023 060 049 029 090 007 008 005 015 086 072 086 044 069 068 099 006
011 095 043 073 058 028 093 097 037 092 001 027 047 088 014 089 063 015 058 056 010
074 007 080 088 038 089 010 099 034 028 072 014 061 046 038 061 094 022 063 040 058
003 093 050 053 016 055 065 081 060 021 039 040 037 046 091 004 018 076 001 100 031
038 043 027 075 004 090 059 025 029 069 024 032 059 052 019 045 094 094 047 063 087
041 079 039 085 020 070 020 042 030 065 033 077 046 066 078 070 093 025 054 068 071
089 035 025 081 085 048 060 097 012 092 053 044 045 042 051 022 036 049 081 033 031
062 043 048 032 080 036 095 091 056 034 074 055 064 091 011 077 077 052 052 005 021
045 099 037 097 044 078 001 079 074 062 097 091 035 004 024 036 028 051 083 085 027

We could have calculated the average happiness level using all 100 respondents. However, this would take us more time (although it is easily doable) than sampling 15 individuals and calculating the average happiness. The formula below is how you calculate a mean (use the box below for your computation):

$$\overline{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$$

With the arithmetic mean you can calculate now the standard deviation as follow:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$



In turn, with the standard deviation we can calculate the standard error:

$$s_{\overline{x}} = \frac{s}{\sqrt{n}}$$

Finally we can use the arithmetic mean to create the sampling distribution (of the mean). If we were to repeat the process above 10 times (i.e., if we constructed 10 samples of 15 individuals each) and graph their means (10 means) we would get a graph that might not look "normally distributed." However, if we repeated this 1,000 times the graph will look "more normal" (and the more samples you draw, the "more normal" your distribution would look like). However, remember that we created a sample of 15 individuals; we probably need a bigger sample size (about 25) to obtain a "normal" distribution. There are online applications that recreate this process, but if you have time you can try it yourself "manually."

Table 1: Happiness in Scotland, 2012

ID	Happiness	ID	Happiness	ID	Happiness
01	8	**41**	7	**81**	8
02	9	**42**	8	**82**	7
03	6	**43**	9	**83**	8
04	7	**44**	10	**84**	9
05	5	**45**	8	**85**	6
06	10	**46**	5	**86**	7
07	9	**47**	9	**87**	6
08	8	**48**	6	**88**	7
09	8	**49**	8	**89**	10
10	1	**50**	9	**90**	10
11	7	**51**	7	**91**	9
12	4	**52**	8	**92**	7
13	8	**53**	10	**93**	7
14	7	**54**	9	**94**	10
15	8	**55**	8	**95**	9
16	10	**56**	8	**96**	8
17	8	**57**	9	**97**	8
18	8	**58**	9	**98**	5
19	9	**59**	8	**99**	8
20	8	**60**	7	**100**	8
21	9	**61**	9		
22	9	**62**	10		
23	8	**63**	7		
24	8	**64**	8		
25	8	**65**	9		
26	8	**66**	10		
27	9	**67**	9		
28	10	**68**	10		
29	8	**69**	9		
30	9	**70**	8		
31	7	**71**	6		
32	9	**72**	10		
33	8	**73**	10		
34	9	**74**	8		
35	4	**75**	10		
36	8	**76**	10		
37	7	**77**	8		
38	9	**78**	8		
39	6	**79**	9		
40	7	**80**	5		