Harald Schmidbauer
Sema Bayraktar
Yonca Demİr

İstanbul Bilgi Üniversitesi

Bus 274: Further Statistics for Business

Spring 2015

PROBLEM SHEET # 3

Problem: In January 2014, the research institute Infratest dimap conducted a survey among eligible voters in Germany. One of the aims of the survey, which had been commissioned by ARD (German public radio and television broadcasters), was to investigate the attitude of German voters toward the immigration from other EU countries to Germany. One of the statements in the survey was:

> "Our economy depends on qualified workforce from other countries
> in order to remain successful."

Respondents were asked if they tend to agree with this statement ("Yes" / "No"). In a representative sample of 1004 voters, 68% said "Yes". Define $p =$ share among all eligible voters in Germany who answer "Yes" when asked this question in January 2014. (The link to the survey results is here.)

a) What does it mean when we speak of a *representative sample*? How could this be obtained in practice? Could we simply ask people in the street on a workday, 10 AM? Why is it important to have a *representative* sample?

b) Which urn model do we use to represent the population of eligible voters in Germany?

c) Sampling from a population corresponds to sampling from the urn. Which stochastic model do we use for the latter

   - if only one element is sampled,
   - if several elements are sampled?

   What exactly is random?

d) We don't know p (why not?), but since there are two possible outcomes we might set $p = 50\%$. Will this reflect the voters' attitude? Why or why not?

e) If 68% in the sample say "Yes", what statement (conclusion) can be made about the population? Is this conclusion deductive or inductive? Why? Is it risky?

f) Now assume that p is actually 0.5. Is it still possible that 68% in a sample of 1004 respondents answer "Yes"?

g) Again assuming $p = 0.5$, compute the probability that *at least* 68% in a random sample of 1004 answer "Yes". (Use the CLT.)

h) In view of the result in (g), is "$p = 0.5$" a safe assumption? Give reasons for your answer.