

## EXAMPLES ON BIAS

**QUESTION 1** FOX News host and commentator Lou Dobbs frequently addresses the issue of illegal immigration, and the viewership of his show tends towards older White males. One of his programs was largely devoted to attacking a proposal to offer driver's licenses to illegal immigrants in California. During the show, Mr. Dobbs invited his viewers to go to [loudobbs.com](http://loudobbs.com) to vote on the question: "Would you be more or less likely to vote for a presidential candidate who supports giving drivers' licenses to illegal aliens?" The result [announced at the end of the show]: 97% of the 7350 people who voted said, "Less likely." [!]

- a) Identify the population of interest, and the sample. Identify and describe the various biases associated with the sampling and explain. **Be thoughtful!**
- b) How should the polling / sampling be done to reduce the incidence of bias identified in a)? [1 sentence]

### **Solution.**

- a) The population of interest is all adult Americans while the sample (likely) consists of predominantly older White Americans.
- b) There is a clear case of sampling or selection bias [under-coverage, specifically] since certain subsets of the population are likely excluded whereas others, richer and Older White Males were likely over-represented since that is the profile of the show's viewers who were then asked to access Mr. Dobb's site...which would require a certain amount of wealth [access to TV, access to internet]. Those responses are likely to systematically differ from those that don't watch Mr. Dobb's show and / or are relatively poor. Further, there is a case Voluntary Response, so that only those that feel strongly and passionately about illegal immigration are likely to make their viewpoints known, and which likely differ from those who aren't as motivated. Bottom-line, this wasn't a representative sample from a biased non-random sampling method.
- b) A random sampling design – any one! – shall reduce the bias that this poll is victim to.

**QUESTION 2** In 1936, *Literary Digest* magazine conducted the most extensive (to that date) public opinion poll in history. They mailed out questionnaires to 10million people whose names and addresses they obtained from telephone books and vehicle registration lists. More than 2.4million people responded, with 57% indicating that they would vote for Republican Alf Landon in the upcoming Presidential election. Incumbent Democrat Franklin Roosevelt won the election, carrying 63% of the popular vote. Identify the population of interest, and identify the

sample. Then, identify and describe the numerous biases in the sampling method; specifically, comment on why the sampling method made it vulnerable to over-estimating support for the Republican candidate. Write 4-5 sentences. **Caution!** This is not a Voluntary Response situation since the respondents did not volunteer themselves for the sample.

## **Solution.**

The population of interest is all American voters or all registered voters or all adult Americans, whereas the sample was only 2.4million Americans who responded to the questionnaires **OR** 2.4million Americans who were sourced from whose names and addresses they obtained from telephone books and vehicle registration lists.

Since the questionnaires were only mailed to individuals sourced from **or** whose names were obtained from telephone books and vehicle registration lists, those individuals would tend to have been wealthier, and more Republican, than the rest of the population. Ergo, there is a **sampling or selection bias – under-coverage**, specifically – since Americans without telephones and / or cars i.e. those of lower income groups...who vote Democratic [!] would have been systematically excluded! There is also a case of **Non-response bias** with a 24% response-rate if those that responded to the poll systematically differed from those that chose not to respond.

**QUESTION 3** Based on a survey conducted on the Diet Smart.com web site, investigators concluded that women who regularly watch Oprah were only one-seventh as likely to crave fattening foods as those who watched other daytime talk shows (*San Luis Obispo Tribune*, October 14, 2000).

a) Identify the population of interest, and identify the sample.

b) Based on a) and the description of the problem, identify 2 likely sources of bias and describe them in detail. **Use the following phrases and terminology while describing bias [whenever applicable!]:**

- non-random sample → sample not representative of the population
- subset [mention which] of the population excluded
- [definition of bias] systematically favouring certain outcomes [mention which!]
- selection / sampling bias; measurement / response bias; non-response bias; convenience sample / sampling; voluntary response sample

## **Solution.**

- a) The population is all women / all dieting women while the sample is only the women that responded to the survey.
- b) Since this is an online survey, the study is prey to sampling / selection bias since it excludes the subset of the population without access to computers / internet. Further, the responses were likely voluntary so that only the most passionate / motivated amongst users would have responded. In both cases, the opinions / characteristics of the respondents might systematically differ from those excluded from the sample, thereby favouring certain outcomes. Ergo, the non-random sampling makes the results unrepresentative. Finally, [Oprah] viewers may be more sensitive to issues of weight, so there *might* be a case of Response bias, too, if they tended to be less candid!

**QUESTION 4** Based on a survey conducted on the Diet Smart.com web site, investigators concluded that women who regularly watch Oprah were only one-seventh as likely to crave fattening foods as those who watched other daytime talk shows (*San Luis Obispo Tribune*, October 14, 2000).

a) Identify the population of interest, and identify the sample.

b) Based on a) and the description of the problem, identify 2 likely sources of bias and describe them in detail. **Use the following phrases and terminology while describing bias [whenever applicable!]:**

- non-random sample → sample not representative of the population
- subset [mention which] of the population excluded
- [definition of bias] systematically favouring certain outcomes [mention which!]
- selection / sampling bias; measurement / response bias; non-response bias; convenience sample / sampling; voluntary response sample

## **Solution.**

a) The population is all women / all dieting women while the sample is only the women that responded to the survey.

b) Since this is an online survey, the study is prey to sampling / selection bias since it excludes the subset of the population without access to computers / internet. Further, the responses were likely voluntary so that only the most passionate / motivated amongst users would have responded. In both cases, the opinions / characteristics of the respondents might systematically differ from those excluded from the sample, thereby favouring certain outcomes. Ergo, the non-random sampling makes the results unrepresentative. Finally, [Oprah] viewers may be more sensitive to issues of weight, so there *might* be a case of Response bias, too, if they tended to be less candid!

