

Case Study 8.1 Effects of Question Wording in Surveys

Kalton et al. [1978]¹ conducted an experiment to assess the effects of wording survey questions in various ways. Part of the investigation looked at whether one gets results, when just one side of a case is presented in a survey question, that are noticeably different from the results obtained when both sides of the case are presented. The experiment was performed in two locations, Lancashire and London. In both locations, one group of people was asked “Are you in favor of giving special priority to buses in the rush hour *or not?*” Another roughly equally sized group was asked “Are you in favor of giving special priority to buses in the rush hour *or should cars have just as much priority as buses?*”² The data presented in Table 1 give the proportion of people agreeing that priority should be given to buses. In each case the number immediately following the proportion in brackets is the size of the sample involved. Because none of the sample proportions are small, and the sample sizes are all around 500, we know that the 10% rule will be satisfied when making comparisons.

Table 1 : Question Effects: Traffic Question^a

	Without reference to cars		With reference to cars		Difference $\hat{d} = \hat{p}_1 - \hat{p}_2$
	\hat{p}_1	(n_1)	\hat{p}_2	(n_2)	
All respondents	0.69	(1,076)	0.55	(1,081)	0.14
Lancashire	0.62	(482)	0.40	(496)	0.22
London	0.76	(594)	0.69	(585)	0.07
Women	0.65	(585)	0.49	(590)	0.16
Men	0.74	(491)	0.66	(488)	0.08
Non Car-owners	0.73	(565)	0.55	(554)	0.18
Car owners	0.66	(509)	0.54	(522)	0.12

^aSource: Kalton et al. [1978].

If p_1 is the underlying true proportion who would agree to priority for buses using the first form of the question which has no reference to cars, and p_2 is the underlying true proportion who would agree when there is a reference to cars, then the difference $d = p_1 - p_2$ measures the effect of the change in the wording of the question. Estimates of the question effect are therefore given

¹Kalton, G. and Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society, Series A*, **145**, 42–73.

²Our italics in both cases are to highlight the differences between the questions.

by the difference column above ($\hat{d} = \hat{p}_1 - \hat{p}_2$). We will calculate question effects for different groups of respondents. In each case, \hat{p}_1 and \hat{p}_2 will come from separate samples of people (cf. Fig. 8.5.1(a) in the text) so that the standard error formula in Table 8.5.5(a) of the text is appropriate.

For the whole sample (“all respondents”), $\hat{p}_1 = 0.69$, $n_1 = 1,076$, $\hat{p}_2 = 0.55$, and $n_2 = 1,081$. The estimated question effect is $\hat{p}_1 - \hat{p}_2 = 0.14$ with standard error

$$\begin{aligned} \text{se}(\hat{p}_1 - \hat{p}_2) &= \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \\ &= \sqrt{\frac{0.69 \times 0.31}{1076} + \frac{0.55 \times 0.45}{1081}} = 0.02068203. \end{aligned}$$

A 95% confidence interval for the true difference is given by

$$\hat{p}_1 - \hat{p}_2 \pm z \text{se}(\hat{p}_1 - \hat{p}_2) = 0.14 \pm 1.96 \times 0.02068203 = [0.10, 0.18].$$

We can say with 95% confidence that, in the population that these people were sampled from, changing the question from the first form to the second form (making reference to cars) would reduce the percentage of people who would agree to giving priority to buses by between 10 and 18 percentage points.

How stable is the question effect we have just seen? To what extent does it differ between people living in Lancashire and London, or between women and men, car owners and non car-owners? We first compare Lancashire and London. It looks as though the question effect (the difference) is larger in Lancashire. For those sampled in Lancashire, $\hat{p}_1 = 0.62$, $n_1 = 482$, $\hat{p}_2 = 0.40$ and $n_2 = 496$. The estimated question effect is $\hat{d}_{Lanc} = \hat{p}_1 - \hat{p}_2 = 0.22$ which has standard error $\text{se}(\hat{d}_{Lanc}) = \text{se}(\hat{p}_1 - \hat{p}_2) = 0.03118762$. This gives a 95% confidence interval for the true question effect, d_{Lanc} , of [0.16, 0.28].

For London respondents, $\hat{p}_1 = 0.76$, $n_1 = 594$, $\hat{p}_2 = 0.69$ and $n_2 = 585$. The estimated question effect is $\hat{d}_{Lond} = \hat{p}_1 - \hat{p}_2 = 0.07$ with $\text{se}(\hat{d}_{Lond}) = \text{se}(\hat{p}_1 - \hat{p}_2) = 0.02593669$ and a 95% confidence interval for the true difference, d_{Lond} , of [0.02, 0.12]. The Lancashire interval ([0.16, 0.28]) and the London interval do not overlap so we have confidence at least 95% that the question effect in Lancashire is larger than that in London, but by how much? Our analysis will now go beyond differences in proportions, but we will find that we still have all the tools we need.

The true difference between the question effects in Lancashire and that in London, namely $d_{Lanc} - d_{Lond}$, is estimated by $\hat{d}_{Lanc} - \hat{d}_{Lond} = 0.22 - 0.07 = 0.15$. These estimates are independent (as they come from physically

independent samples) so that the standard error of the difference is given by (see Section 7.5 in the text)

$$\begin{aligned} \text{se}(\hat{d}_{Lanc} - \hat{d}_{Lond}) &= \sqrt{\text{se}(\hat{d}_{Lanc})^2 + \text{se}(\hat{d}_{Lond})^2} \\ &= \sqrt{0.03118762^2 + 0.02593669^2} = 0.04056328. \end{aligned}$$

The estimates are also approximately Normally distributed so that a 95% confidence interval for the difference in question effects is given by $0.15 \pm 1.96 \times 0.04056328$ or between 0.07 and 0.23. The effect of changing the question would be bigger in Lancashire by between 7 and 23 percentage points than it would be in London.

How could a difference like this occur? Not only is there a considerably smaller question effect in London, but there is also more support for bus lanes. It could be that Londoners have had more experience with bus lanes and many are already aware that giving priority to buses means that cars have less priority. This would mean that that they would be more likely to respond the same way when this consequence of priority for buses is spelled out as they would when it is not spelled out. Differences between the relative use made of private cars and buses between the two centers could also be a factor.

We then went on to investigate the extent to which gender makes a difference to question effect, and whether car ownership makes a difference (see 95% confidence intervals in Table 2).

Table 2 : Question Effects

Factor	95% CIs for true values					
Location	Lanc:	[0.16, 0.28]	Lond:	[0.02, 0.12]	Diff:	[0.07, 0.23]
Gender	Women:	[0.10, 0.22]	Men:	[0.02, 0.14]	Diff:	[0.00, 0.16]
Own a car?	No:	[0.13, 0.24]	Yes:	[0.06, 0.18]	Diff:	[-0.02, 0.14]

We cannot say with 95% confidence that gender makes a difference to the question effect, although it could be larger for women by up to 16 percentage points. Similarly, we cannot say whether owning a car makes a difference to the question effect. Let us lower the level of confidence to 90%. At the 90% level of confidence we can say that the question effect is bigger for women than for men by between 1 and 15 percentage points. Even at the 90% confidence level, the interval for the difference between the question effect for car owners and the question effect for non car-owners contains zero (90% CI is [-0.01, 0.13]).

In the same study, the researchers also compared the questions “Do you think that side streets should be closed to through traffic *or not*?” and “Do

you think that side streets should be closed to through traffic *or should through traffic be free to take any route it likes?*”. In this case they could not detect any question effect. One explanation is that the alternative to the proposition about closing side streets to through traffic was obvious to most respondents.