

Sampling: Final and Initial Sample Size Determination



Chapter Outline

- 1) Overview
- 2) Definitions and Symbols
- 3) The Sampling Distribution
- 4) Statistical Approaches to Determining Sample Size
- 5) Confidence Intervals
 - i. Sample Size Determination: Means
 - ii. Sample Size Determination: Proportions
- 6) Multiple Characteristics and Parameters
- 7) Other Probability Sampling Techniques

Chapter Outline

8) Adjusting the Statistically Determined Sample Size

9) Non-response Issues in Sampling

i. Improving the Response Rates

ii. Adjusting for Non-response

10) International Marketing Research

11) Ethics in Marketing Research

12) Summary

Definitions and Symbols

- **Parameter:** A **parameter** is a summary description of a fixed characteristic or measure of the target population. A parameter denotes the true value which would be obtained if a census rather than a sample was undertaken.
- **Statistic:** A **statistic** is a summary description of a characteristic or measure of the sample. The sample statistic is used as an estimate of the population parameter.
- **Finite Population Correction:** The **finite population correction** (fpc) is a correction for overestimation of the variance of a population parameter, e.g., a mean or proportion, when the sample size is 10% or more of the population size.

Definitions and Symbols

- **Precision level:** When estimating a population parameter by using a sample statistic, the **precision level** is the desired size of the estimating interval. This is the maximum permissible difference between the sample statistic and the population parameter.
- **Confidence interval:** The **confidence interval** is the range into which the true population parameter will fall, assuming a given level of confidence.
- **Confidence level:** The **confidence level** is the probability that a confidence interval will include the population parameter.

Symbols for Population and Sample Variables

Table 12.1

Variable	Population	Sample
Mean	μ	\bar{X}
Proportion	Π	p
Variance	σ^2	s^2
Standard deviation	σ	s
Size	N	n
Standard error of the mean	$\sigma_{\bar{X}}$	$S_{\bar{X}}$
Standard error of the proportion	σ_p	S_p
Standardized variate (z)	$(X-\mu)/\sigma$	$(X-\bar{X})/S$
Coefficient of variation (C)	σ/μ	s/\bar{X}

The Confidence Interval Approach

Calculation of the confidence interval involves determining a distance below (\bar{X}_L) and above (\bar{X}_U) the population mean (\bar{X}), which contains a specified area of the normal curve (Figure 12.1).

The z values corresponding to \bar{X}_L and \bar{X}_U may be calculated as

$$z_L = \frac{\bar{X}_L - \mu}{\sigma_{\bar{x}}}$$

$$z_U = \frac{\bar{X}_U - \mu}{\sigma_{\bar{x}}}$$

where $z_L = -z$ and $z_U = +z$. Therefore, the lower value of \bar{X} is

$$\bar{X}_L = \mu - z\sigma_{\bar{x}}$$

and the upper value of \bar{X} is

$$\bar{X}_U = \mu + z\sigma_{\bar{x}}$$

The Confidence Interval Approach

Note that μ is estimated by \bar{X} . The confidence interval is given by

$$\bar{X} \pm z\sigma_{\bar{x}}$$

We can now set a 95% confidence interval around the sample mean of \$182. As a first step, we compute the standard error of the mean:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 55/\sqrt{300} = 3.18$$

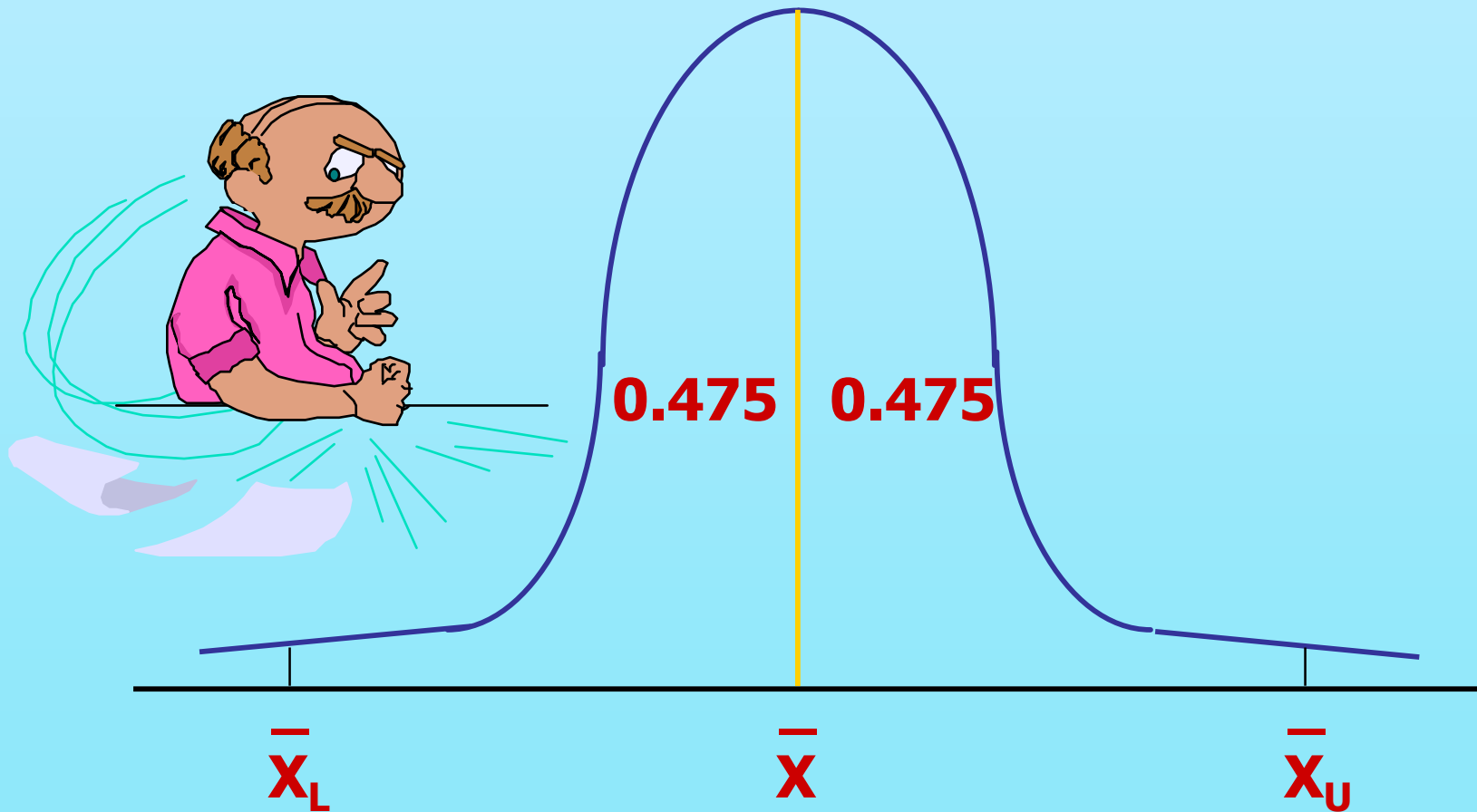
From Table 2 in the Appendix of Statistical Tables, it can be seen that the central 95% of the normal distribution lies within ± 1.96 z values. The 95% confidence interval is given by

$$\begin{aligned} & \bar{X} \pm 1.96\sigma_{\bar{x}} \\ & = 182.00 \pm 1.96(3.18) \\ & = 182.00 \pm 6.23 \end{aligned}$$

Thus the 95% confidence interval ranges from \$175.77 to \$188.23. The probability of finding the true population mean to be within \$175.77 and \$188.23 is 95%.

95% Confidence Interval

Figure 12.1



Sample Size Determination for Means and Proportions

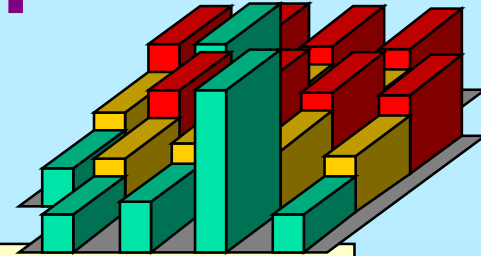



Table 12.2

Steps	Means	Proportions
1. Specify the level of precision	$D = \pm\$5.00$	$D = p - \Pi = \pm 0.05$
2. Specify the confidence level (CL)	CL = 95%	CL = 95%
3. Determine the z value associated with CL	z value is 1.96	z value is 1.96
4. Determine the standard deviation of the population	Estimate σ : $\sigma = 55$	Estimate Π : $\Pi = 0.64$
5. Determine the sample size using the formula for the standard error	$n = \sigma^2 z^2 / D^2 = 465$	$n = \Pi(1-\Pi) z^2 / D^2 = 355$
6. If the sample size represents 10% of the population, apply the finite population correction	$n_c = nN / (N+n-1)$	$n_c = nN / (N+n-1)$
7. If necessary, reestimate the confidence interval by employing s to estimate σ	$= \bar{X} \pm zS_x$	$= p \pm zS_p$
8. If precision is specified in relative rather than absolute terms, determine the sample size by substituting for D.	$D = R\mu$ $n = C^2 z^2 / R^2$	$D = R\Pi$ $n = z^2(1-\Pi) / (R^2\Pi)$

Sample Size for Estimating Multiple Parameters

Table 12.3

	Variable		
	Mean Household Department store shopping	Monthly Expense On Clothes	Gifts
Confidence level	95%	95%	95%
z value	1.96	1.96	1.96
Precision level (D)	\$5	\$5	\$4
Standard deviation of the population (σ)	\$55	\$40	\$30
Required sample size (n)	465	246	217



Adjusting the Statistically Determined Sample Size

Incidence rate refers to the rate of occurrence or the percentage, of persons eligible to participate in the study.

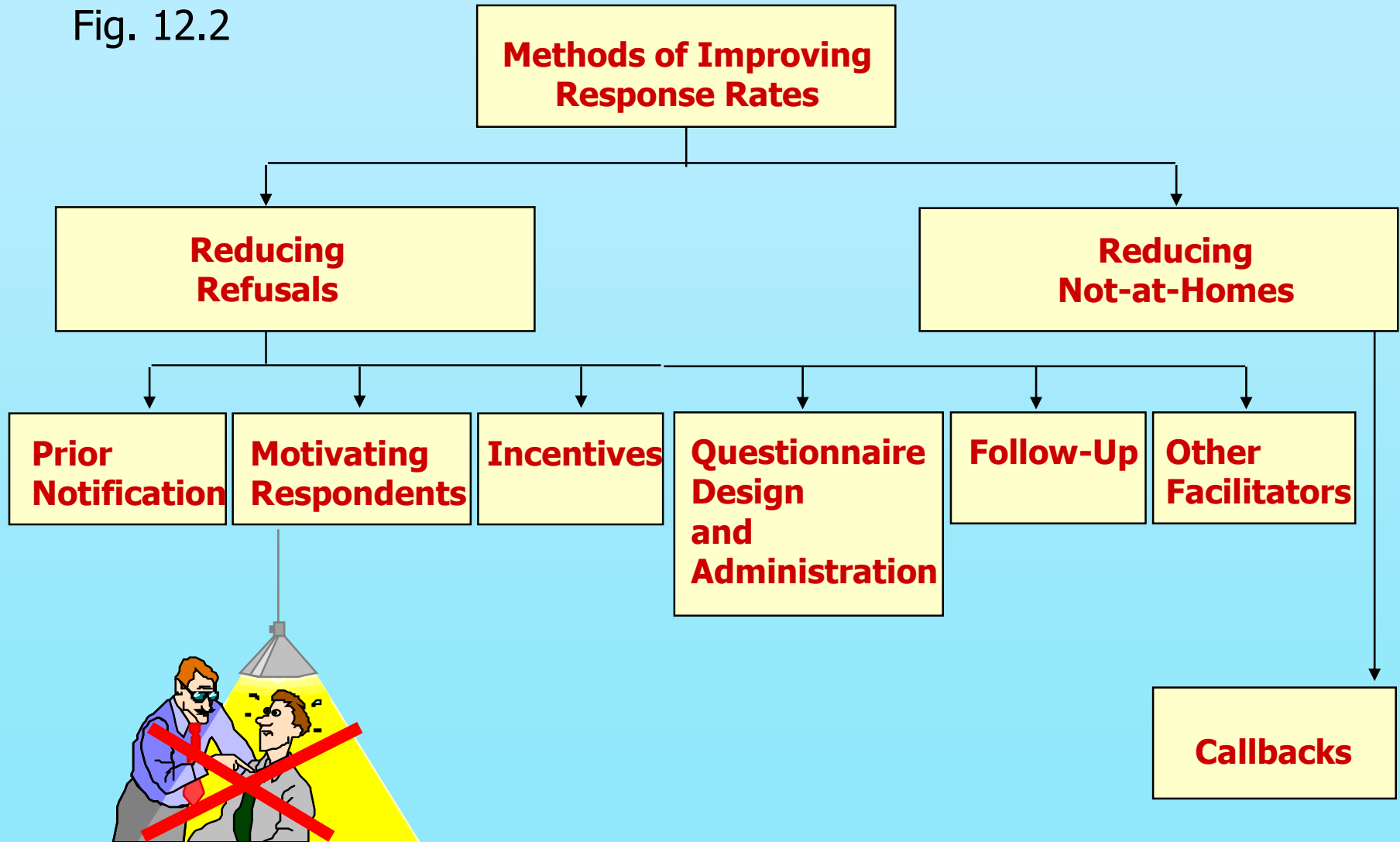
In general, if there are c qualifying factors with an incidence of $Q_1, Q_2, Q_3, \dots, Q_c$, each expressed as a proportion:

$$\text{Incidence rate} = Q_1 \times Q_2 \times Q_3 \dots \times Q_c$$

$$\text{Initial sample size} = \frac{\text{Final sample size}}{\text{Incidence rate} \times \text{Completion rate}}$$

Improving Response Rates

Fig. 12.2



Arbitron Responds to Low Response Rates

Arbitron, a major marketing research supplier, was trying to improve response rates in order to get more meaningful results from its surveys. Arbitron created a special cross-functional team of employees to work on the response rate problem. Their method was named the “breakthrough method,” and the whole Arbitron system concerning the response rates was put in question and changed. The team suggested six major strategies for improving response rates:

1. Maximize the effectiveness of placement/follow-up calls.
2. Make materials more appealing and easy to complete.
3. Increase Arbitron name awareness.
4. Improve survey participant rewards.
5. Optimize the arrival of respondent materials.
6. Increase usability of returned diaries.



Eighty initiatives were launched to implement these six strategies. As a result, response rates improved significantly. However, in spite of those encouraging results, people at Arbitron remain very cautious. They know that they are not done yet and that it is an everyday fight to keep those response rates high.

Adjusting for Nonresponse

- **Subsampling of Nonrespondents** – the researcher contacts a subsample of the nonrespondents, usually by means of telephone or personal interviews.
- In **replacement**, the nonrespondents in the current survey are replaced with nonrespondents from an earlier, similar survey. The researcher attempts to contact these nonrespondents from the earlier survey and administer the current survey questionnaire to them, possibly by offering a suitable incentive.

Adjusting for Nonresponse

- In **substitution**, the researcher substitutes for nonrespondents other elements from the sampling frame that are expected to respond. The sampling frame is divided into subgroups that are internally homogeneous in terms of respondent characteristics but heterogeneous in terms of response rates. These subgroups are then used to identify substitutes who are similar to particular nonrespondents but dissimilar to respondents already in the sample.

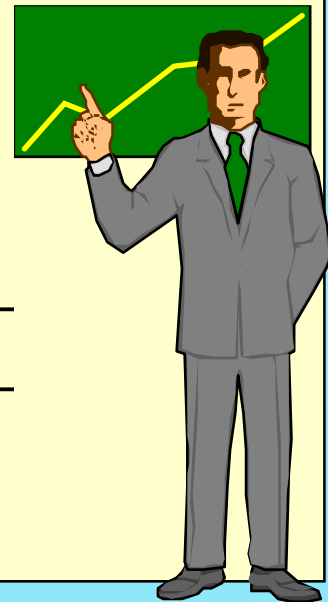
Adjusting for Nonresponse

- **Subjective Estimates** – When it is no longer feasible to increase the response rate by subsampling, replacement, or substitution, it may be possible to arrive at subjective estimates of the nature and effect of nonresponse bias. This involves evaluating the likely effects of nonresponse based on experience and available information.
- **Trend analysis** is an attempt to discern a trend between early and late respondents. This trend is projected to nonrespondents to estimate where they stand on the characteristic of interest.

Use of Trend Analysis in Adjusting for Nonresponse

Table 12.4

	Percentage Response	Average Dollar Expenditure	Percentage of Previous Wave's Response
First Mailing	12	412	—
Second Mailing	18	325	79
Third Mailing	13	277	85
Nonresponse	(57)	(230)	91
Total	100	275	



Adjusting for Nonresponse

- **Weighting** attempts to account for nonresponse by assigning differential weights to the data depending on the response rates. For example, in a survey the response rates were 85, 70, and 40%, respectively, for the high-, medium-, and low income groups. In analyzing the data, these subgroups are assigned weights inversely proportional to their response rates. That is, the weights assigned would be $(100/85)$, $(100/70)$, and $(100/40)$, respectively, for the high-, medium-, and low-income groups.

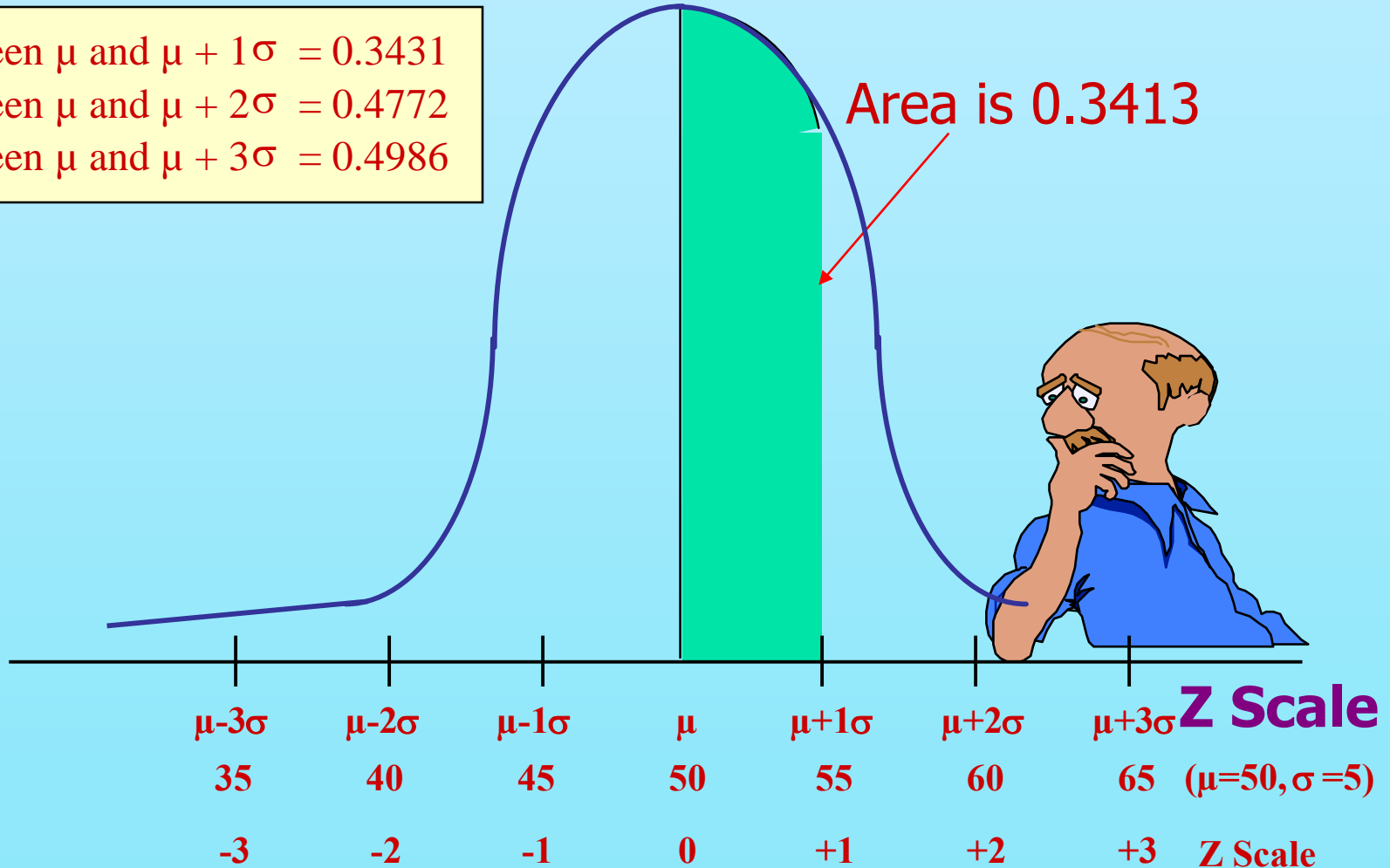
Adjusting for Nonresponse

- **Imputation** involves imputing, or assigning, the characteristic of interest to the nonrespondents based on the similarity of the variables available for both nonrespondents and respondents. For example, a respondent who does not report brand usage may be imputed the usage of a respondent with similar demographic characteristics.

Finding Probabilities Corresponding to Known Values

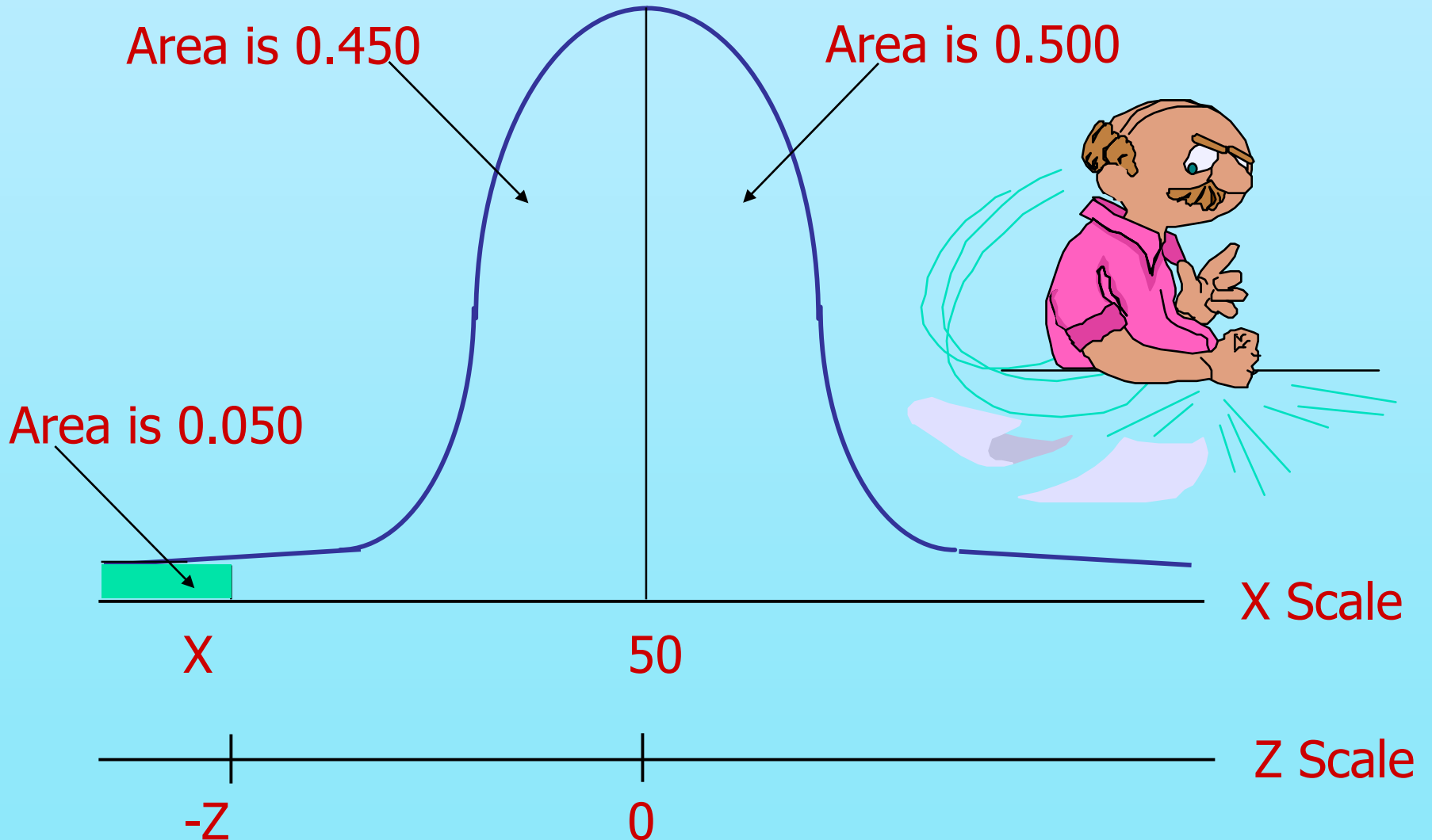
Figure 12A.1

Area between μ and $\mu + 1\sigma = 0.3431$
Area between μ and $\mu + 2\sigma = 0.4772$
Area between μ and $\mu + 3\sigma = 0.4986$



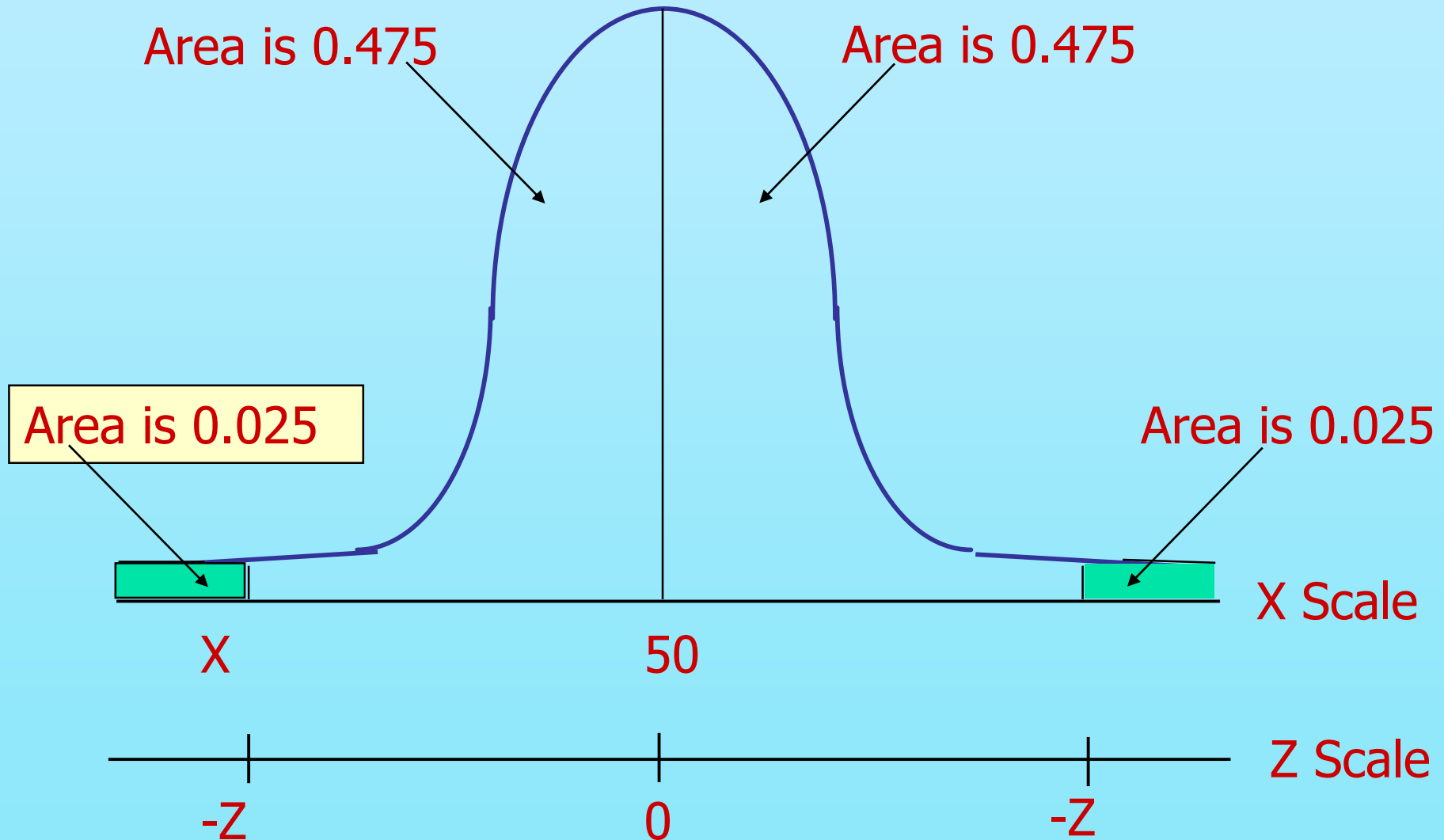
Finding Probabilities Corresponding to Known Values

Figure 12A.2



Finding Values Corresponding to Known Probabilities: Confidence Interval

Fig. 12A.3



Opinion Place Bases Its Opinions on 1000 Respondents

Marketing research firms are now turning to the Web to conduct online research. Recently, four leading market research companies (ASI Market Research, Custom Research, Inc., M/A/R/C Research, and Roper Search Worldwide) partnered with Digital Marketing Services (DMS), Dallas, to conduct custom research on AOL.

DMS and AOL will conduct online surveys on AOL's *Opinion Place*, with an average base of 1,000 respondents by survey. This sample size was determined based on statistical considerations as well as sample sizes used in similar research conducted by traditional methods. AOL will give reward points (that can be traded in for prizes) to respondents. Users will not have to submit their e-mail addresses. The surveys will help measure response to advertisers' online campaigns. The primary objective of this research is to gauge consumers' attitudes and other subjective information that can help media buyers plan their campaigns.

Opinion Place Bases Its Opinions on 1000 Respondents

Another advantage of online surveys is that you are sure to reach your target (sample control) and that they are quicker to turn around than traditional surveys like mall intercepts or in-home interviews. They also are cheaper (DMS charges \$20,000 for an online survey, while it costs between \$30,000 and \$40,000 to conduct a mall-intercept survey of 1,000 respondents).

