Tutorial 3

T3 – 1. The purpose of this practical Tutorial 3 is to illustrate the practice and theory of equiprobable selecting. This Tutorial is to be done *individually*, not in groups.

The respondent population is 100 plastic blocks of various sizes and shapes; the response of interest is block *weight*. The population attribute to be estimated is the *average* weight of the population of blocks, using a sample of size 10 obtained by:

(a) *judgement* selecting; (b) *equiprobable* selecting.

- First, choose 10 blocks that you judge to be representative of the respondent (or study) population; record their numbers and then look up their weights on the overhead display.
- Second, choose 10 two-digit equiprobable numbers and then look up the weights of the blocks with these numbers.

Use the two *upper* tables provided below to record your observations. When you have the information for your two samples, calculate and record it (in appropriate units) in the two smaller *lower* tables near the bottom of this page 0.29-1:

– the *sample average* ($\bar{y}$);
– the *sample (data) standard deviation* (s);
– the *estimated standard deviation of the sample average* (equiprobable selecting only);
– *approximate 90% and 95% confidence intervals for* $\bar{\mathbf{Y}}$ (equiprobable selecting only).

Have your calculations completed as soon as possible so that the results of everyone in the class can be combined on displays that the instructor will construct on overhead transparencies (and you will copy).

Judgement Selecting	
Block number	**Block weight**

Equiprobable Selecting	
Block number	**Block weight**

$\bar{y}$	
s	

$\bar{y}$	
s	
$\hat{s.d.}(\bar{Y})$	
90% CI for $\bar{\mathbf{Y}}$	
95% CI for $\bar{\mathbf{Y}}$	

(*continued overleaf*)

T3 – 1. (cont.) The following questions are provided to stimulate group discussion and individual study of the statistical ideas illustrated by this Tutorial 3.

1 Refer to the two bar-graphs of sample *averages* on the third side (page 0.26-3) of the Tutorial.

- What matters from the theory of equiprobable selecting for estimating $\overline{\mathbf{Y}}$ (the respondent population *average*) are illustrated by the bar-graphs, with particular reference to:
  - the *mean* of $\overline{Y}$;
  - the *standard deviation* of $\overline{Y}$;
  - the *distribution* of $\overline{Y}$;
  - modelling the sample average under repeated equiprobable selecting as a *random variable*?
- What are the essential *differences* between the bar-graphs for samples selected by equiprobable and judgement selecting?
  - What are the *practical* implications for sample surveys of these differences?

2 Refer to the two bar-graphs of sample *standard deviations* on the fifth side (page 0.26-5) of the Tutorial.

- What matters from the theory of equiprobable selecting for estimating $\overline{\mathbf{Y}}$ (the respondent population *average*) are illustrated by the bar-graphs, with particular reference to:
  - the *mean* of S;
  - the *standard deviation* of S;
  - the *distribution* of S;
  - modelling the sample standard deviation under repeated equiprobable selecting as a *random variable*?
- What are the essential *differences* between the bar-graphs for samples obtained by equiprobable and judgement selecting?
  - What are the *practical* implications for sample surveys of these differences?
- Compare and contrast your answers to Question 1 for sample *averages* with Question 2 for sample *standard deviations*.

3 In the context of the expressions under equiprobable selecting for $s.d.(\overline{Y})$ [which involves $\mathbf{S}$] and $\widehat{s.d.}(\overline{Y})$ [which involves S as an *estimator* of $\mathbf{S}$], consider the following two matters:

- the lower bar-graph on the fifth side (page 0.26-5) of the Tutorial. suggests that S as an estimator of $\mathbf{S}$ has *substantial* variability under repeated equiprobable selecting;
- comparison of tables of the standard normal and t distributions shows that there are only *small* differences in the corresponding critical values for degrees of freedom (or samples sizes) as small as even about 10.

Explain how these matters can be reconciled with the expressions (involving the standard normal and t_{n-1} distributions, respectively) for a confidence interval for $\overline{\mathbf{Y}}$ (the respondent population average) when $\mathbf{S}$ [the respondent population (data) standard deviation] is *known* and when it is *un*known and is estimated by s [the sample (data) standard deviation].

4 Refer to the two diagrams (on the fourth and sixth sides (pages 0.26-4 and 0.26-6) of the Tutorial showing 90% and 95% confidence intervals for $\overline{\mathbf{Y}}$ (the respondent population average) under repeated equiprobable selecting.

- Describe *in words* what is meant by a $100(1-\alpha)\%$ confidence interval for $\overline{\mathbf{Y}}$ (the respondent population average).
- Indicate how the two confidence interval diagrams, which are based on the class results shown in the lower bar-graphs on the third and fifth sides (pages 0.26-3 and 0.26-5) of the Tutorial, illustrate the main points in your description.
- Describe, in point form, the *essential* differences between the diagrams for confidence levels of 90% and 95%.
- Discuss critically the following statements:
  - *An advantage of interval estimating over point estimating is that it provides a quantitative indication of* both *precision* and *accuracy*.
  - *The size of the sample (1,446 Canadian citizens 18 years old and over in the 10 provinces) is expected to yield a reflection of all Canadian voters to an accuracy of plus or minus 2.6 percentage points 19 times out of 20.*
  - *The probability that the true value of the population average lies within a correctly-calculated 99% confidence interval is about 0.99.*

5 A bar-graph of the *population* of 100 block weights shows that the distribution of weights is *moderately skewed* to the right. Discuss the implications of this matter for the interpretation of the confidence intervals calculated from samples of 10 blocks obtained by equiprobable selecting.

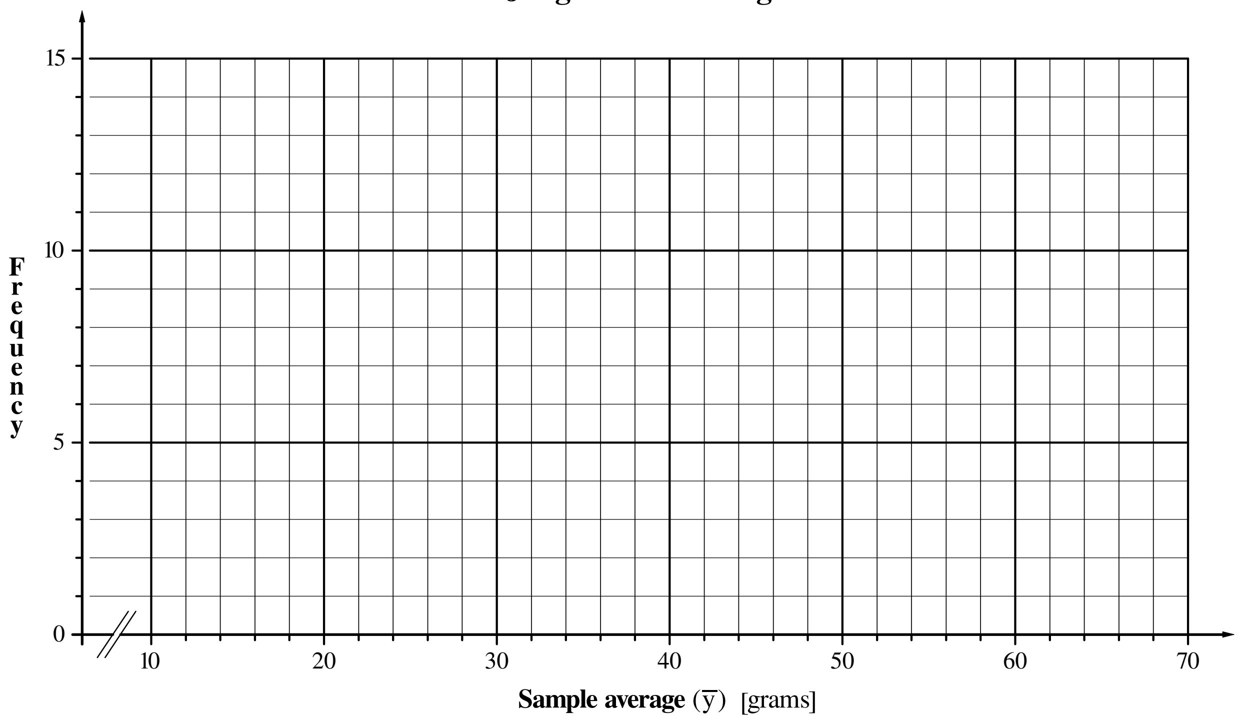
6 Outline the *statistical* reason(s) why a confidence interval should *not* be calculated on the basis of the responses of a set of respondent population units obtained by *judgement* selecting.

(*continued*)

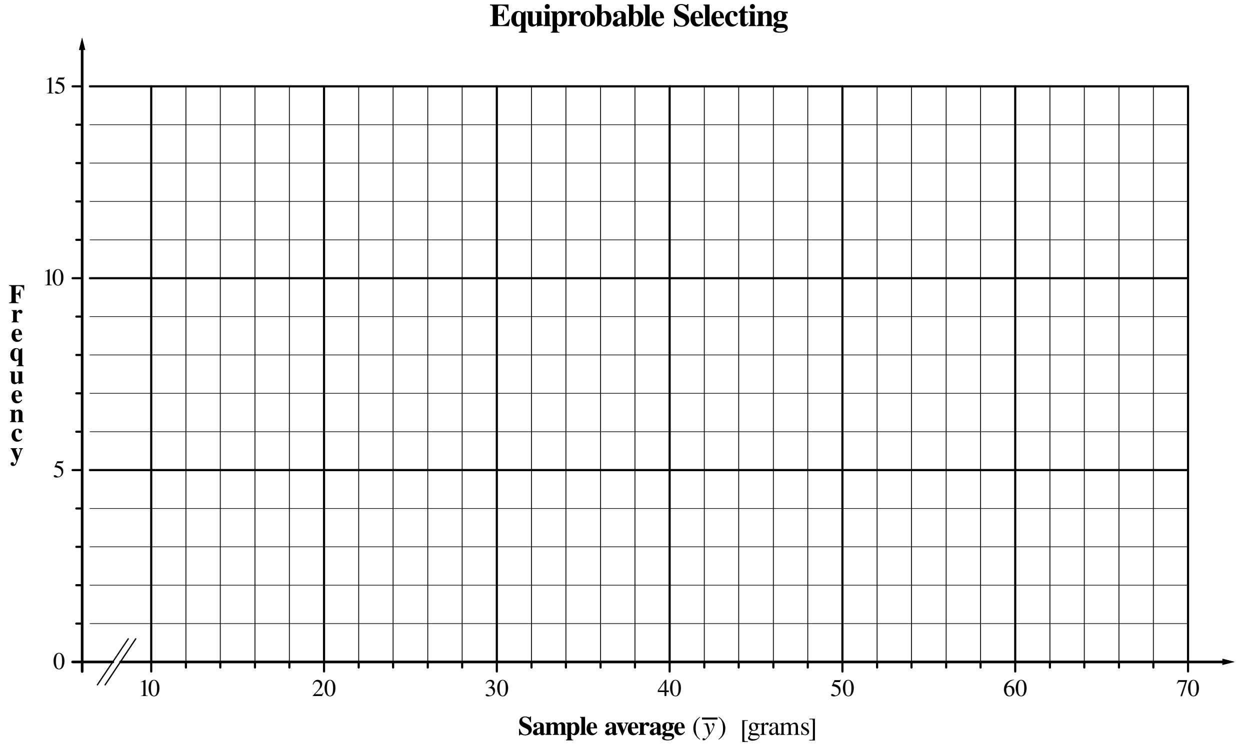
Tutorial 3 (continued 1)

ESTIMATING $\overline{\mathrm{Y}}$ – Bar-graphs of sample averages

Judgement Selecting



Equiprobable Selecting

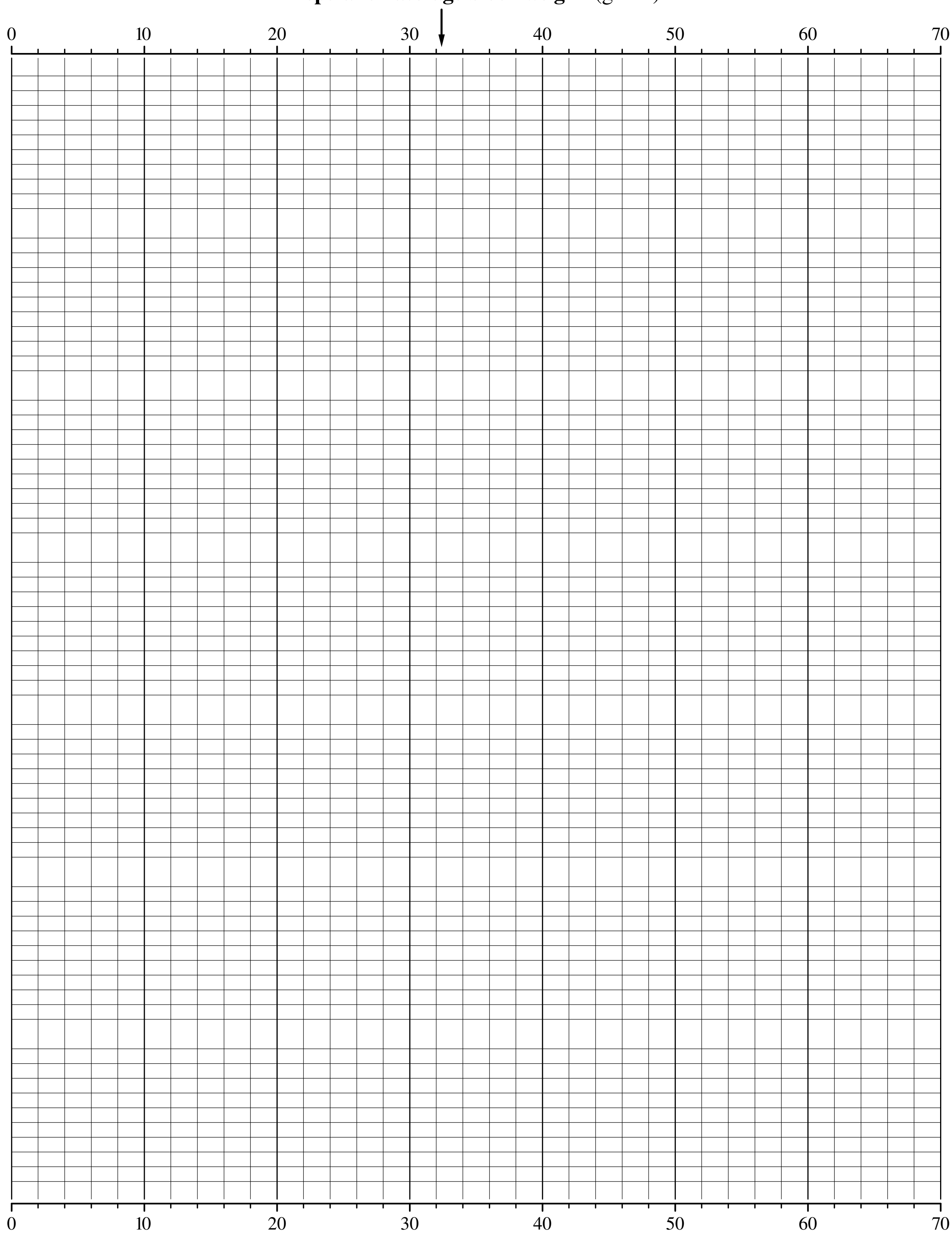


(*continued overleaf*)

T3 – 1.
(cont.)

ESTIMATING $\overline{Y}$ – 90% Confidence intervals from repeated equiprobable selecting

Population average block weight (grams)



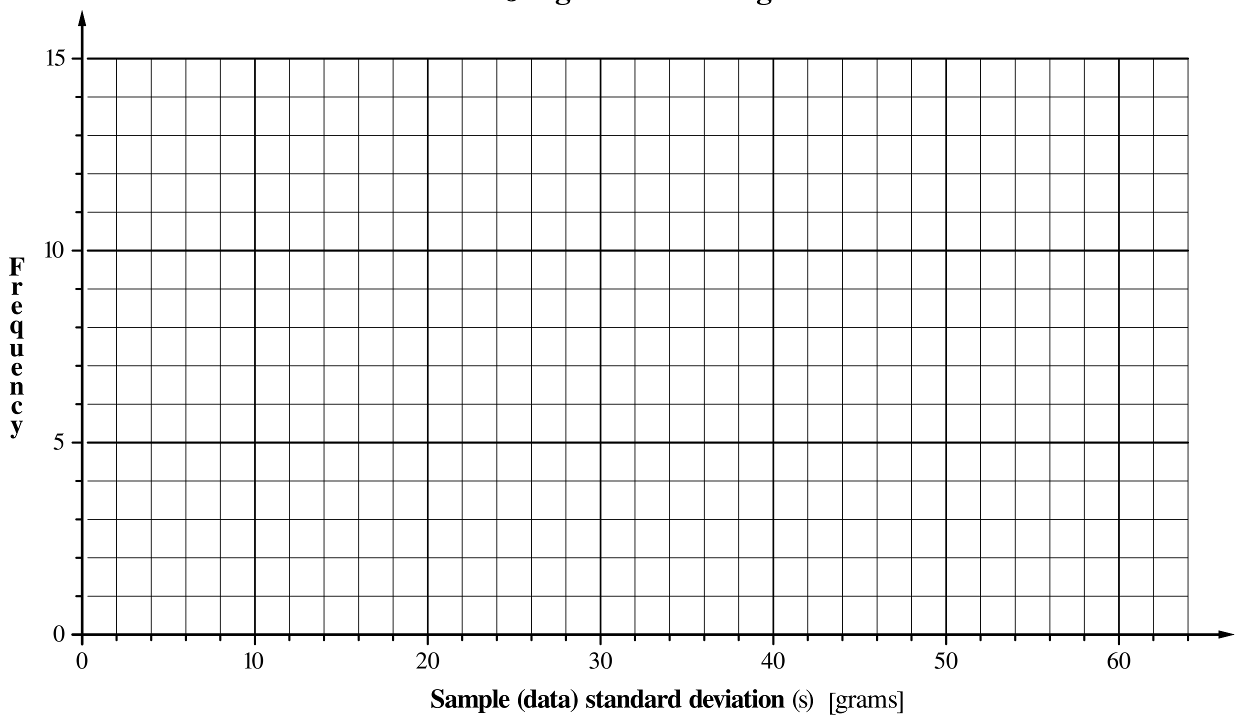
Population average block weight (grams)

(*continued*)

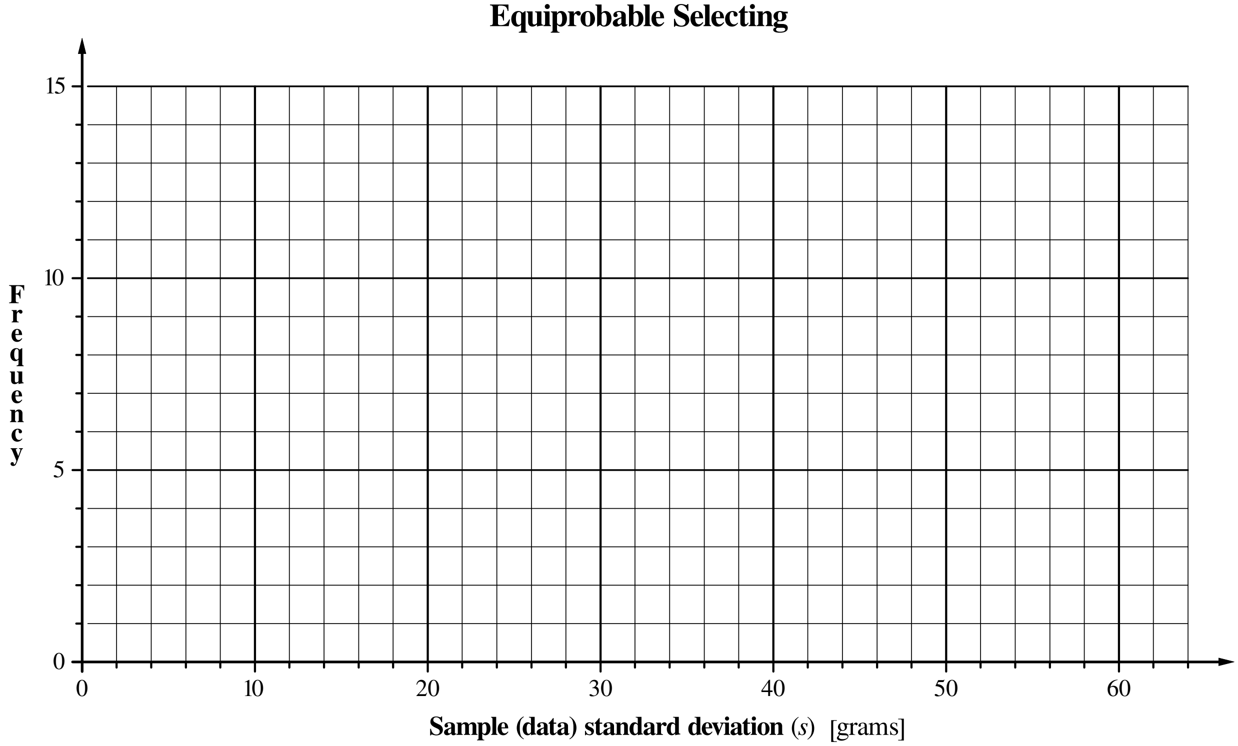
Tutorial 3 (continued 2)

ESTIMATING ~~S~~ – Bar-graphs of sample (data) standard deviations

Judgement Selecting



Equiprobable Selecting



(*continued overleaf*)

T3 – 1.
(cont.)

ESTIMATING $\overline{\mathrm{Y}}$ – 95% Confidence intervals from repeated equiprobable selecting