Emergency Food Security Assessments (EFSAs)
Technical guidance sheet n°. 11

Using the T-square sampling method to estimate population size, demographics and other characteristics in emergency food security assessments (EFSAs)[1]

Table of Content

[1] Alden Henderson, consultant, WFP Food Security Analysis Service, December 2008.

Using the T-square sampling method to estimate population size, demographics and other characteristics in emergency food security assessments (EFSAs)

This Technical Guidance Sheet (TGS) describes the T-square (T2) sampling method, and provides guidance and instructions on when and how to use the method in an EFSA. The T2 is a sampling design that can be used instead of other sampling methods in an EFSA survey. The advantage of the T2 method is that it can estimate the population size and the numbers of people with certain characteristics, such as food insecurity status, livelihood or age, without constructing a sampling frame. This TGS complements other guidance for EFSAs in the *EFSA Handbook*.[2], in particular Part IV which deals with estimating numbers of people at livelihood risk and projecting numbers of people in need of assistance.

1. Why do population size and demographics change in a crisis?

A shock such as a tsunami, drought or armed conflict may damage roads, crops and livestock, homes, markets and medical clinics. This destruction creates hardship for the people who live or work in the affected area. As a result, they may leave for areas that are safer and that offer better opportunities for food, income, shelter and security. These movements often change the population size and demographics of both the abandoned and the resettled areas. Accurate and timely information on the numbers and locations of people, and on their needs, is a key element of an effective and appropriate humanitarian response.

2. What are the key steps in the T-square method?

The T2 method described in this TGS is a modified version of a method used in ecology to determine the spatial distribution and density of trees and plants in large areas. In this method, the number of trees is calculated by measuring the distance from a random point to its closest tree, or “nearest neighbour”. The distance to the nearest neighbour is related to density and, when multiplied by the total area, yields the total number of trees in an area. This TGS modifies the ecology method to estimate human population size and demographics by measuring the distance between a random point and an occupied house and counting the number of occupants in the house. Section 4 describes how to apply the modified T2 method. The key steps are:

a. selecting a random point;
b. measuring the distance from this point to the nearest occupied house;
c. going to this house, finding the occupied house nearest to it, and measuring the distance between the two houses;
d. counting the number of people living in the two houses and calculating the average;
e. estimating the total area of the survey;
f. calculating the total number of houses in the survey area by dividing the total area of the survey by the average space occupied by each house and its surroundings; and
g. calculating the population size by multiplying the number of houses in the survey area by the average number of people in a household.

When an EFSA survey examines specific characteristics of households, such as their food insecurity situation or reliance on specific livelihood activities, these results can be used to determine the numbers of people with each characteristic by multiplying the total population size by the percentage of people with the characteristic.

---

[2] See revised *EFSA Handbook,* Part IV Section 3, Conducting a situation analysis, WFP Food Security Analysis Service, 2009.

3. When is it appropriate to use the T square method?

The following questions provide a guideline for when to use the T2. If the answers to all the following questions are "yes", the T2 sampling method can be used. If the answers to questions a to d are "yes", but any answer to questions e to h is "no", the T2 should be considered with the other sampling methods suggested, to determine which is most appropriate.

a) Is a sampling frame unavailable, unreliable or outdated, and would it be difficult, expensive or laborious to construct one?

If YES, the T2 method may be appropriate.
If NO, consider using a random, systematic, proportional or cluster sample.

b) Is the sampling frame complex?

If YES, the T2 method can simplify the sample design by reducing the number of stages. This may also reduce the design effect, thereby decreasing the sample size necessary to obtain the necessary precision.
If NO, consider using a random, proportional or cluster sample.

c) Is the survey taking place during the recovery stage of a crisis?

If YES, the T2 method may be appropriate if there is enough time. It takes one or two weeks to identify and train surveyors in the method before starting the survey.
IF NO, other methods require less time and fewer resources, and would be more appropriate in the immediate post-impact phase of an acute emergency. These include the Delphi method, aerial surveys, cluster surveys, quadrant surveys and using available census and survey data to estimate population size and demographics. WFP has TGS for cluster and quadrant surveys**,** and the Delphi method.[3]

d) Is the area where the survey will be conducted larger than 10 km^2?

If YES, the T2 method may be appropriate.
If NO, consider a random, proportional, cluster or quadrant sample.

e) Are most of the houses in the survey area single-storey residences?

If YES, the T2 method may be appropriate.
If NO, consider a proportional or cluster sample.

f) Are most of the houses randomly distributed?

If YES, the T2 method may be appropriate.
If NO, a modified T2 method (Annex I) may be appropriate,
but requires more time and resources.

g) Are detailed maps available?

If YES, the T2 method may be appropriate.
If NO, use a cluster sample design to select households randomly.

---

[3] See Technical Guidance Sheets No. 7 *Area Method to Estimate Population Size and Demographics in Emergency Food Security Assessments (EFSAs),* A. Henderson, WFP Emergency Needs Assessment Service (now Food Security Analysis Service), September 2007; and No. 10. *Using the Delphi Method to Estimate Population Size and Demographics in Emergency Food Security Assessments (EFSAs),* A. Henderson, WFP Food Security Analysis Service, January 2008.

The T2 method can be used to find the random starting point for each cluster.

h) Are global positioning system (GPS) units available to survey teams?

If YES, the T2 method may be appropriate.
If NO, use a cluster sample design to select households randomly, and the T2 method to find which houses to survey.

4. What are the activities for the T-square sampling procedure?

The T2 method produces three numbers that are used to calculate the population size in the target area:

1) the total size of the survey area;
2) the average land area occupied by a household; and
3) the average number of people in a household.

The T2 method can be divided into pre-field, field and post-field activities. The main pre-field activities are identifying the area to survey and deciding the level of precision, or margin of error, for the estimate. Field activities include recruiting and training the survey teams, conducting the T2 sampling procedure and ensuring that responses are accurately recorded. Post-field activities involve analyzing and interpreting the data and reporting the results.

4.1 - Pre-field activities

Once it has been decided to use the T2 sampling method in the ESFA survey, and administrative and logistics support are available, the next steps are to define the area to be surveyed, decide on the margin of error for the survey, calculate the number of households to sample, and select and mark random points on the map.

a. ***Defining the area to be surveyed***
This is usually the area where the shock has occurred, people are located and assistance will be provided. Geographical Information System (GIS)[4] programs are used to identify the survey area's borders and calculate its surface area. If GIS is not available, Google Earth (http://earth.google.com/) or local maps can be used to estimate the area to be surveyed.

A refinement of identifying the target area is to include only the inhabited areas in the survey area. This can be accomplished by including areas around cities and villages and excluding uninhabited areas such as rivers, lakes, deserts, ravines, mountains, swamps, parks, forests, etc. This will decrease the amount of area the survey will cover and the variability of the estimate. If the uninhabited areas cannot be identified, alternate approaches are to:

1) Have GIS programs identify buffers around geographical features associated with habitation, such as roads etc., and position random points in these buffered areas.

2) Increase the number of random points to be surveyed to account for uninhabited areas. Local staff can define what constitutes an uninhabited area, such as when there is no occupied house within 1 km of a random point (the horizon is about 1.7 km away for a person 2 m tall standing in a flat area), or no house within a 30-minute walk from the random point. The survey team records this area as uninhabited and moves to the next random point. The number of points surveyed should be increased to account for uninhabited areas.

b. ***Deciding the margin of error/level of precision for the survey***
The margin of error/level of precision for the population estimate obtained from the survey can be:

[4] Among other agencies, the Food Security Analysis Service at WFP Headquarters may be able to provide assistance with this.

- ± 20 to 30 percent for surveys done ten to 15 days after the shock; or
- ± 10 percent (or less) for surveys done 15 days after the shock.

The margin of error can be changed according to how the ESFA results are to be used. For instance, if the EFSA results are to be compared with previous or future surveys, a margin of error of less than 10 percent may be more appropriate; the smaller the margin of error, however, the more households have to be surveyed.

c. *Calculating the number of households to be sampled*

The margin of error is used to calculate the number of households to be surveyed, which should be less than 10 percent of the total households in the area. The following table provides a guide to the number of houses to sample according to the precision of the survey and using a simple random sample design. The "estimated percentage of the characteristic in the surveyed population" refers to the population characteristic that the survey is measuring –insecure households, children under 5, farmers, etc. If this percentage is not known, 50 percent should be used as the estimate, because this produces the largest sample size.

	Estimated percentage of the characteristic in the surveyed population									
Margin of error (%)	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**
5	473	724	943	1 131	1 288	1 413	1 507	1 570	1 601	1601
10	160	219	269	313	348	375	395	407	411	407
15	88	113	134	151	165	176	182	186	186	182
20	66	72	82	91	98	103	106	107	106	103

d. Selecting and marking random points on a map of the area to be surveyed

If possible, systematic sample selection should be used to locate random points for the survey. Systematic selection is preferred because it identifies a point in each section of the survey area, so the surveyed households are spread across the entire survey area. A random sample may have clusters of houses in one area, and/or large areas that are not sampled.

4.2 - Field activities

T2 field activities calculate the average area a household occupies and the average household size in the target area. These numbers, along with the total size of the target area, are used to estimate the population size and, if additional details are collected, demographics. The following steps are carried out in the field:

a. ***Recruiting survey teams*****:** When possible, recruit people who have experience of conducting surveys and know how to administer a questionnaire and record responses. As a minimum, each team should have one leader and one assistant. The team leader navigates to the random point, finds the nearest neighbour and house T (see Figure 1), explains the purpose of the survey to the household occupants and obtains their approval for conducting the survey. The assistant measures the distances from the random point to its nearest neighbour, and from there to the next nearest neighbour. Both administer the questionnaire to household T.

b. ***Training surveyors:*** Team members need special training on conducting the T2 method. People take time to grasp the concept of finding the nearest neighbour and consistently measuring the distance between the houses. One day of training – a morning session to explain the concept and an afternoon session to practise the T2 sample procedure in the field, followed by a discussion on any problems encountered – helps team members to become comfortable with and proficient in this step.

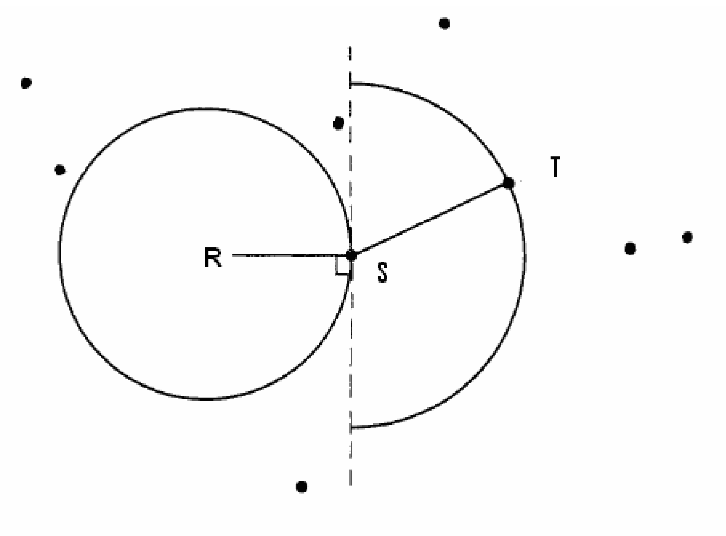
c. ***Obtaining maps and equipment for the survey*****:** Survey teams need detailed maps showing the survey area, each random point and nearby landmarks so that they can find the boundaries of the survey area and navigate to the random points. WFP Headquarters, regional bureaux or country offices may be able to provide these maps. If not, they can be downloaded from Google Earth (http://earth.google.com/, Figure 3). Methods for measuring the distances between houses are also needed. These can be as simple as training survey teams to measure their paces and count the number of paces to reach a house, or to use a tape or rope marked in metres. Specialized measuring tools such as a surveyor's wheel or a laser range finder speed up the measuring process.

d. ***Navigating to the random point (R):*** There are two ways of reaching the random point, depending on whether or not GPS instruments are available:

1. When GPS instruments are available: GIS software marks the random points (R on Figure 1) and landmarks in the survey area (Figure 2), calculates the distances and directions from each landmark to each random point, and provides the latitude and longitude of each random point and landmark. The survey team inputs these numbers into a GPS instrument and uses it to navigate to a random point. This is best done by going to a nearby landmark and using GPS to navigate to the random point. Once at the random point, the survey team conducts the T2 sampling as illustrated in Figure 1.

2. When GPS instruments are not available: Teams mark random points on locally available maps or those drawn by GIS or Google Earth, and identify the locations of nearby landmarks (Figure 3). The survey team uses the map to navigate to the random point, where it conducts the T2 sampling as illustrated in Figure 1.

Figure 1. Using the T2 method to identify which houses to survey



Source: adapted from Diggle, 1982.

Each dot in Figure 1 is a house. The survey team uses GPS instruments or maps to navigate to the random point, R. Once at R, the team visually locates the nearest household (S) and goes to it. At S, the team looks for the next nearest household (T) within the hemisphere that is perpendicular to the line drawn from R to S. The survey team measures the distances from R to S and from S to T, and interviews the members of household T. The distance from R to S is x and the distance from S to T is z in the formulas used to calculate population size and margin of error.

e. ***Taking measurements using the T2 sampling method*****:** Use Figure 1 and the following instructions:

1. Once at R, the team marks it with a coloured flag or stake. It then locates the nearest house (S) by visual observation, and walks to it. When two or more households appear to be at similar distances, the team counts the number of paces it takes to reach each, to identify which is the closest. If two or more

houses are equidistant from R, the one to be surveyed is selected randomly. The team then measures the distance from R to S.

2. At S, the team determines whether the house is occupied or vacant. A vacant house is usually one where no one has lived for the past three months, but survey teams can use a different reference period based on local events or conditions. If S is vacant, the team goes back to R and selects the next nearest house, repeating these steps until the nearest occupied house is identified.
3. A team member stands in front of the occupied house S, faces R and extends her/his arms outwards. The angle between the outstretched arms and the line from R to S forms the T (Figure 1), but none of the houses inside this T can be selected as house T. The team member turns 180º – one of the houses in front of him/her can be selected as house T. The team looks for the nearest occupied house – house T – and measures the distance from S to T. The survey is administered to occupants of house T.
4. The random point is completed when the team has: (1) measured the distance from point R to house S; (2) measured the distance from house S to house T; and (3) interviewed the occupants of house T.
5. The T2 sample procedure is designed for areas where each building is inhabited by one household. Other combinations include several families sharing a multi-storey building, compound of several huts, or single hut. Typically, a household is defined as a group of people who eat from the same pot and sleep under the same roof. The shock may have changed living conditions so that several families are living in one household, but if they all eat and sleep under the same roof, they are considered to be one family.

f. ***Identifying household respondents and obtaining demographic and other information:*** As in any survey, the team leader must explain the purpose of the survey to potential participants, provide any necessary information and ask permission to administer the questionnaire. If a participant refuses, surveyors should ask how many people live in the household, or obtain the information from a neighbour.

4.3 Post-field activities

At this stage, the data are analyzed and a report produced on the estimated population size and its margin of error. As mentioned earlier, the T2 sampling procedure produces three numbers for calculating the population size in the target area:

1) the total size of the survey area;

2) the average land area occupied by a household; and

3) the average number of people in a household.

The total surface area surveyed divided by the average area occupied by a household indicates how many houses are in the survey area. Multiplying this number by the average number of people in a household gives the population size of the survey area. The following formula is used:

$$\frac{total \cdot size \cdot of \cdot survey \cdot area}{average \cdot area \cdot of \cdot household} \times \begin{matrix} average \cdot number \cdot of \\ people \cdot in \cdot household \end{matrix} = \begin{matrix} population \cdot in \\ survey \cdot area \end{matrix}$$

F1

The average land area a house occupies includes the land it sits on, the land associated with it – garden, animal pen, paths/roads, uninhabited structures such as latrines, storage, animal housing, etc. – and the land between it and the next house.

As illustrated in the following example, the report should include at least the estimated average number of people in the survey area, and the margin of error and range of the estimate. Additional elements can be the number of houses sampled, how many were vacant, and any problems encountered during the survey that may have affected the results and population estimates.

The accompanying Excel spreadsheet contains formulas that automatically calculate the population size and confidence interval when survey team members input the distances between houses and the numbers of occupants.

Example

The T2 method determined the following:

a. Average land area occupied by a household: 150 m^2.
b. Average number of people in a household: 7.6.
c. Total size of target area: 1.25 million m^2.

Total population: 1,250,000 m^2
------------------ X 7.6 people/house = (8,333.3 houses x 7.6 people/house) = 63,330 people
150 m^2

Final results: The total number of people in the survey area is estimated at 63,330. This estimate has a 10 percent margin of error, and the true population will be between 57,000 and 69,660.

5. What are the limitations of the T-square sampling procedure?

a) The nearest neighbour to a random point on the edge of the target area may be outside the survey area, and teams must select only houses that are within the area. In this situation, the nearest neighbour in the survey will not be the house closest to the random point, so the nearest neighbour distance may be overestimated. The survey team should note that the true nearest neighbour is outside the survey area, and that it has used the nearest house within that area.

b) Surveyors must understand the method and be able to find point R and houses S and T, measure distances accurately and use the same point of reference for each house.

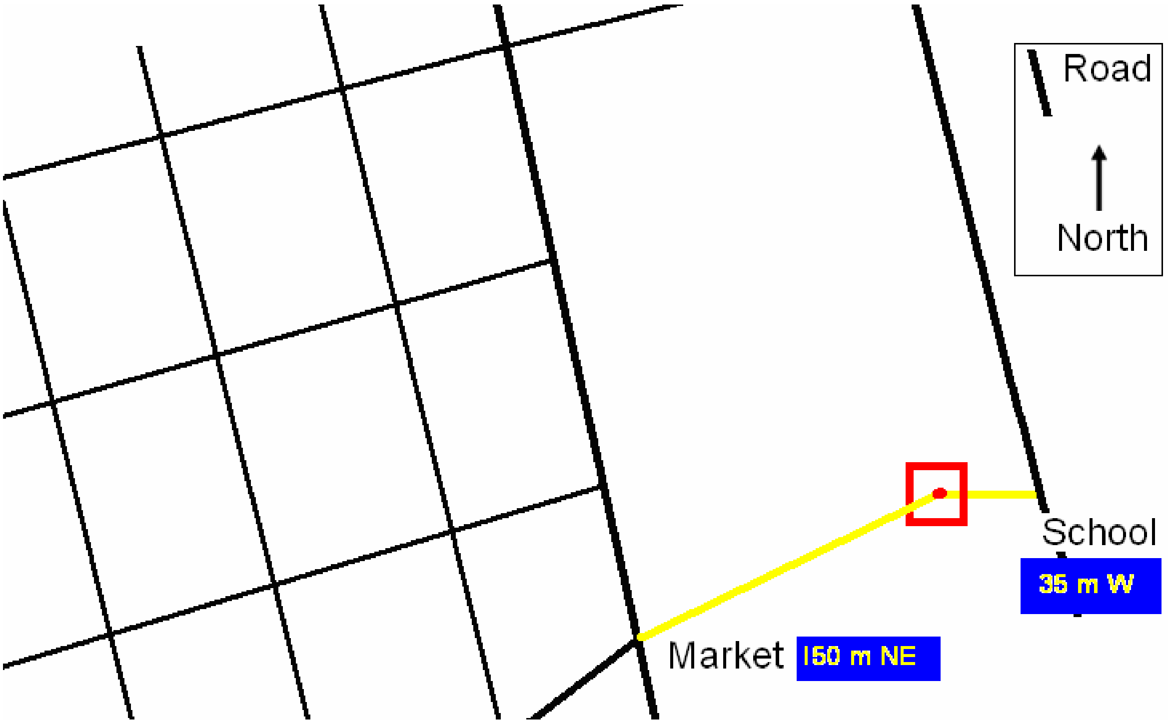
c) The reference point used to measure the distance from one house to another may vary when the shapes and sizes of houses vary. Survey teams should agree what reference point in the house to use – for example, the front door, a specific corner of the house, or the midpoint of a specific wall on the house. At the end of each day, survey teams should discuss situations where they had to make decisions. The dimensions of houses can be measured and reported to show the variation in house shape.

d) The T2 method is designed to estimate the density of households, so it is assumed that the households in the sample area are not clustered. If there is clustering, the alternative T2 method described in Annex I should be used.

e) As for any survey method applied in the field, before using the T2 method, it is important to consider whether insecurity, long distances and access to survey areas will hinder survey teams' ability to travel to survey sites. If so, alternative methods should be considered.

Figure 2. Sample map for navigating to a random point

Detail of area surrounding waypoint 16 16' 59.50" N, 95 40' 57.03". Nearby landmarks of a market and a school are identified with their distances and directions to the waypoint.



References

Byth, K. and Ripley, B.D. 1980. On Sampling Spatial Patterns by Distance Methods. *Biometrics,* 36: 279–284.

Diggle, P. 1982. Some Statistical Aspects of Spatial Distribution Models for Plants and Trees. *Studia Forestalia Suecia,* I: 162.

Greenwood, J.D. 1996. Basic techniques. *In* W.J. Sutherland, ed. *Ecological census techniques: a handbook,* pp. 11–110. Cambridge University Press, Cambridge, UK.

Krebs, C.J. 1999. *Ecological Methodology.* 2nd edition. Chapter 5. Estimating Abundance:Line Transects and Distance Methods

These Technical Guidance Sheets, the EFSA Handbook
and other related resources are available at:

www.wfp.org/food-security

Annex I.

Alternative T2 method for use in areas with clustering of households

This procedure is based on the Byth and Ripley procedure for distance methods in large areas with clustering of households; it includes a step that adjusts for the natural clustering of households (Byth and Ripley, 1980). The steps in the procedure are:

1. Double the number of households to survey because two surveys will be carried out and two estimates made: one using the distance from a random point to a house; the other using the distance from a random house to its nearest neighbour.

2. Randomly select half of the starting points. This number is *n*. For example, in Figure 4, the sample is based on three points; doubling this means that six points will be selected. In this case, randomly select three of the six points (marked as unboxed Ps in Figure 4). A survey team goes to each of these three points and measures the distance from it to its nearest occupied household. This is the *x* value. The survey team administers the questionnaire to the occupants of the nearest household. The following formula is used to calculate household density in the survey area:

$$\hat{N}_1 = \frac{n}{\pi\Sigma(x_i^2)} \qquad \text{F2}$$

$\hat{N}_1 =$ estimated population density of households in the survey area;
$n =$ number of points sampled;
$x_i =$ distance from a random point to the nearest household in the survey area.

3. Around each of the remaining three points, mark a small plot containing about five households (shown as boxed Ps in Figure 4). Number all the households in the box and randomly select *n* households from across all the plots. *n* is the number of plots – in this example, three. Because the three houses are randomly selected from all the houses in all the plots, a plot may have no, one, two or even three selected households in it. In Figure 4, one plot contains no selected households, one has one household, and another has two. The survey team measures the distance from the random point to the randomly selected house. This distance is *z*. The team administers the questionnaire to the occupants of the household. The following formula is used to calculate household density in the survey area:

$$\hat{N}_2 = \frac{n}{\pi\Sigma(z_i^2)} \qquad \text{F3}$$

$\hat{N}_2 =$ estimated population density of households in the plots;
$n =$ number of points/plots sampled;
$z_i =$ distance from the random household to its nearest household in the plot.

4. If N_1 and N_2 are statistically similar (use a T-test), the distribution of houses in the survey area is random. The average of N_1 or N_2 can be used as the population estimate in the survey area. If N_1 and N_2 are statistically different, the distribution of the households in the survey area is not random, and formula F4 is used to adjust for the clustering effect:

$$\hat{N}_3 = \sqrt{\hat{N}_1 \hat{N}_2} \qquad \text{F4}$$

$\hat{N}_3 =$ estimated population size when calculating its geometric average from point-to-households within the survey zone $(\hat{N}_1)$ and within the plots $(\hat{N}_2)$.

5. The margin of error for N_1 or N_2 is calculated by:

   defining variance as $\frac{1}{\hat{N}}$, F5

   the variance for each N will be $\frac{\hat{y}^2}{n}$ F6

   and the margin of error for each N will be $\sqrt{\frac{\hat{y}}{n}}$. F7

   $\hat{N}$ is either N_1 or N_2 and *n* is the sample size.

6. The margin of error for $\hat{N}_3$ is calculated by:

   defining variance as $\frac{1}{\hat{N}_3} = \frac{(1/\hat{N}_3)^2}{n}$ F8

   the standard error is: $\frac{1}{\hat{N}_3} = \sqrt{\frac{\text{var}\,iance(1/\hat{N}_3)}{n}}$ F9

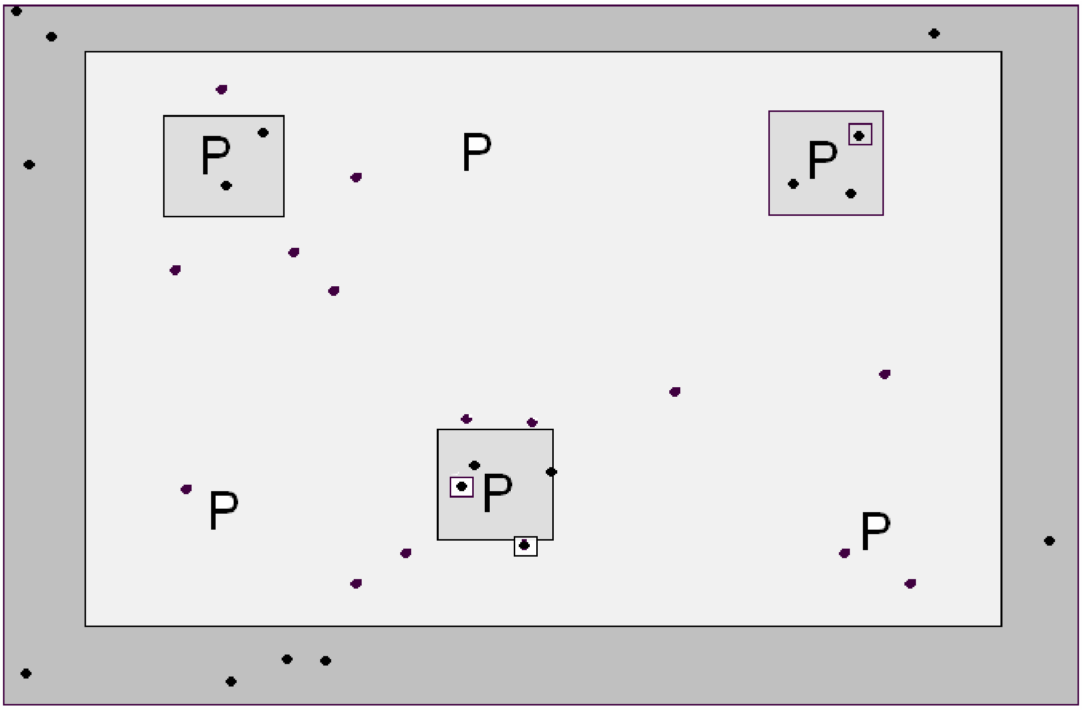
$$\hat{N}_1 = \frac{n}{\pi\Sigma(x_i^2)} \qquad \text{F10}$$

$$\hat{N}_2 = \frac{n}{\pi\Sigma(z_i^2)} \qquad \text{F11}$$

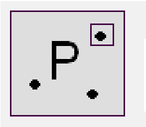
$$\hat{N}_3 = \sqrt{\hat{N}_1 \hat{N}_2} \qquad \text{F12}$$

$(\hat{N}) =$ average area a household occupies;
$(n) =$ sample size;
$x_i =$ distance from the random point to its nearest neighbour;
$z_i =$ distance from the random point to its nearest neighbour in the plot.

Figure 4. Sample map for using the T2 method in a target area with clustering of households



Source: Krebs, 1999.

P = random point;
• = house;
▣ = randomly selected house;
□ = survey area;
■ = boundary strip around survey area;


= plot.

The border of the survey area is marked and a boundary area about 25 m wide is drawn round this boundary. This survey has a sample size (*n*) of 3: 2*n* = (2 x 3) = 6, so six random points are selected and marked systematically in the survey area. Three survey points are randomly selected for the N_1 method (formula F2) and the three remaining points are used for the N_2 method (formula F3). For the N_1 method, a survey team navigates to the random point and finds the distance to its nearest neighbour. For the N_2 method, the team marks a plot containing up to five houses around each random point, numbers each house in the plots consecutively, and randomly chooses *n* houses (three in this case because the sample size is three) from all the plots. Once the houses have been selected, the survey team measures the distance from the random point to randomly selected house(s) in the plot. An example of the data and calculations involved are described in the following:

Table 1.

House number	x_i	x_i^2	z_i	z_i^2	Number of occupants
1	8.65	74.82	3.60	12.96	7
2	12.20	148.84	8.55	73.10	3
3	6.95	48.30	2.15	4.62	5
4	3.05	9.30	6.80	46.24	8
5	9.65	93.12	5.05	25.50	6
6	4.35	18.92	10.60	112.36	4
7	7.10	50.41	4.35	18.92	3
8	15.20	231.04	2.85	8.12	7
9	6.35	40.32	7.95	63.20	6
10	12.00	144.00	3.15	9.92	7
11	2.80	7.84	6.90	47.61	9
12	5.55	30.80	3.95	15.60	2
13	8.10	65.61	8.10	65.61	4
14	11.45	131.10	4.50	20.25	9
15	13.80	190.44	7.65	58.52	8
16	7.35	54.02	1.10	1.21	10
17	6.30	39.69	3.40	11.56	3
18	9.60	92.16	4.80	23.04	7
19	10.35	107.12	6.25	39.06	6
20	3.15	9.92	2.90	8.41	8
Sum	163.95	1 587.80	104.60	665.84	122.00
Average	8.32	81.66	5.22	33.06	6.19
	$\hat{N}_1$	0.004011	40.11		
	$\hat{N}_2$	0.009566	95.66		
	$\hat{N}_3$	0.006195	61.95		

Figures for *x* and *z* were obtained from Kerbs, 1999: Box 5.2.
Number of occupants was randomly selected to centre around 6.

Using the numbers in Table 1, calculation of the population is as follows:

$$\hat{N}_1 = \frac{n}{\pi\Sigma(x_i^2)} = \frac{20}{(3.14)(1587.80)} = 0.004011$$ house/m^2 = 40 houses/ha;

$$\hat{N}_2 = \frac{n}{\pi\Sigma(z_i^2)} = \frac{20}{(3.14)(665.84)} = 0.09566$$ house/m^2 = 96 houses/ha;

$$\hat{N}_3 = \sqrt{\hat{N}_1\hat{N}_2} = \sqrt{(0.004011)(0.09566)} = 0.006195$$ house/m^2 = 62 houses/ha.

As 40 ≠ 96, the houses are clustered in the survey area, and N_3 is a less biased estimate of the population.

Variance = $$\frac{1}{\hat{N}_3} = \frac{(1/\hat{N}_3)^2}{n} = \frac{(1/0.006195)^2}{20} = \frac{(161.5)^2}{20} = 1304.27$$

Standard error = $\frac{1}{\hat{N}_3} = \sqrt{\frac{\text{var}\,iance(1/\hat{N}_3)}{n}} = \sqrt{\frac{1304.27}{20}} = 8.075$

Upper 95% CI = $\frac{1}{N_3} + t_\alpha[S.E.(1/N_3)] = 161.55 + (2.09)(8.075) = 178.429$ m^2/house

= 0.005604 house/m^2

= 56 houses/ha.

Lower 95% CI = $\frac{1}{N_3} - t_\alpha[S.E.(1/N_3)] = 161.55 - (2.09)(8.075) = 144.673$ m^2/house

= 0.006912 house/m^2

= 69 houses/ha.

The 95 percent confidence interval (CI) = 2.09 where t_α for *n* -1 = (20-1) = 19 = 2.09 (see Annex 3).

The population can be estimated from formula F1, using the survey area and average number of people in a household (6.19 in this example). The 95 percent CI can be calculated with an upper and a lower 95 percent CI (69 and 56 in this example).

Annex 2.

Example of data analysis from a T2 sample method that estimates population size and demographics

Data from a survey can be entered in the Excel spreadsheet below to obtain a population estimate.

Table 2. Measurements from a T2 survey

House surveyed	Distance from random point R to house S	Square of distance from R to S	Distance from house S to house T	Square of distance from S to T	Number of people living in house T
1	42	1 764	11	121	7
2	23	529	10	100	3
3	35	1 225	14	196	5
4	32	1 024	18	324	8
5	14	196	21	441	6
6	39	1 521	30	900	4
7	30	900	27	729	3
8	62	3 844	36	1 296	7
9	24	576	38	1 444	6
10	45	2 025	15	225	7
11	17	289	48	2 304	9
12	30	900	20	400	2
13	41	1 681	61	3 721	4
14	45	2 025	27	729	9
15	51	2 601	75	5 625	8
16	23	529	31	961	10
17	52	2 704	86	7 396	3
18	34	1 156	39	1 521	7
19	31	961	100	10 000	6
20	35	1 225	90	8 100	8
Sum	705	27 675	797	46 533	122
Average	35.3		39.9		6.1

x_i = distance from random point R to house S in metres (column 2);

x_i^2 = square of distance from R to S (column 3);

z_i = distance from house S to house T in metres (column 4);

z_i^2 = square of distance from S to T (column 5);

N = number of people living in house T (column 6);
n = sample size (20);

s_x^2 = variance of R to S;

s_{xz} = covariance of the distance from R to S and the distance from S to T.

The density of houses is estimated as:

$$N_T = \frac{n^2}{2\Sigma(x_i)\left[\sqrt{2}\Sigma(z_i)\right]} = \frac{20^2}{2(705)\left[\sqrt{2}(797)\right]} = 0.000366$$ houses/m^2 or 3.7 houses/ha. F2

The standard error is calculated as:

S.E. = $\frac{1}{\hat{N}_T} = \sqrt{\frac{8\left(\bar{z}^2 s_x^2 + 2\bar{x}\bar{z}_{xz} + \bar{x}^2 s_z^2\right)}{n}}$

F3

$$= \sqrt{\frac{8(797)^2(140.2) + 2(705)(797)(77.6) + (705)^2(777.5)}{20}} = 1094$$

$$Variance \cdot of \cdot (x) = \frac{\Sigma x^2 - (\Sigma x)^2 / n}{n-1} = \frac{27675 - (705)^2 / 20}{20-1} = 140.2$$

F4

$$Variance \cdot of \cdot (z) = \frac{\Sigma z^2 - (\Sigma z)^2 / n}{n-1} = \frac{46533 - (797)^2 / 20}{20-1} = 777.5$$

F5

$$CoVariance \cdot of \cdot x \cdot and \cdot z) = \frac{\Sigma xz - (\Sigma x)(\Sigma z) / n}{n-1} = \frac{29569 - (705)(797) / 20}{20-1} = 3.92$$

F6

The 95 percent CI for the reciprocal of this density estimate is:

$\frac{1}{\hat{N}_T} \pm t_\alpha \left[S.E.(\frac{1}{\hat{N}_T} \right]$ where $t_\alpha = 1.711$ for 19 (n-1) degrees of freedom (see Annex 3) F7

$3973 \pm (1.711)(1094) =$ or 2,101 and 5,845

Taking reciprocals gets 0.000476 and 0.000171, which yields 4.76 and 1.71 houses/ha.

If the total area is 10,000 ha and the average number of people in a household is 6.1, the population size is:

10,000 ha x 3.36 x 6.1 = 205,180
10,000 ha x 1.71 x 6.1 = 290,338
10,000 ha x 4.76 x 6.1 = 104,363

Thus, the population size is 205,200 with a 95 percent CI of 104,400 and 220,300. The survey estimated that 16 percent of these people are farmers. To calculate the number of farmers find 16 percent of 205,200 = 24,600.

Annex 3.

Table for determining t_α

		t																		
		Degrees of freedom (df)																	p value	
Confidence	z	100	50	40	30	25	20	15	12	10	9	8	7	6	5	4	3	2	1	2-tailed	1-tailed
0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.500
10%	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.16	.90	.450
20%	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.27	0.27	0.28	0.29	0.32	.80	.400
30%	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.40	0.40	0.40	0.40	0.40	0.41	0.41	0.42	0.44	0.51	.70	.350
40%	0.52	0.53	0.53	0.53	0.53	0.53	0.53	0.54	0.54	0.54	0.54	0.55	0.55	0.55	0.56	0.57	0.58	0.62	0.73	.60	.300
50%	0.67	0.68	0.68	0.68	0.68	0.68	0.69	0.69	0.70	0.70	0.70	0.71	0.71	0.72	0.73	0.74	0.76	0.82	1.00	.50	.250
60%	0.84	0.85	0.85	0.85	0.85	0.86	0.86	0.87	0.87	0.88	0.88	0.89	0.90	0.91	0.92	0.94	0.98	1.06	1.38	.40	.200
70%	1.04	1.04	1.05	1.05	1.05	1.06	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.13	1.16	1.19	1.25	1.39	1.96	.30	.150
80%	1.28	1.29	1.30	1.30	1.31	1.32	1.33	1.34	1.36	1.37	1.38	1.40	1.41	1.44	1.48	1.53	1.64	1.89	3.08	.20	.100
85%	1.44	1.45	1.46	1.47	1.48	1.49	1.50	1.52	1.54	1.56	1.57	1.59	1.62	1.65	1.70	1.78	1.92	2.28	4.17	.15	.075
90%	1.64	1.66	1.68	1.68	1.70	1.71	1.72	1.75	1.78	1.81	1.83	1.86	1.89	1.94	2.02	2.13	2.35	2.92	6.31	.10	.050
91%	1.70	1.71	1.73	1.74	1.75	1.76	1.78	1.81	1.84	1.88	1.90	1.93	1.97	2.02	2.10	2.23	2.47	3.10	7.03	.09	.045
92%	1.75	1.77	1.79	1.80	1.81	1.82	1.84	1.88	1.91	1.95	1.97	2.00	2.05	2.10	2.19	2.33	2.61	3.32	7.92	.08	.040
93%	1.81	1.83	1.85	1.86	1.88	1.89	1.91	1.95	1.99	2.03	2.06	2.09	2.14	2.20	2.30	2.46	2.76	3.58	9.06	.07	.035
94%	1.88	1.90	1.92	1.94	1.95	1.97	1.99	2.03	2.08	2.12	2.15	2.19	2.24	2.31	2.42	2.60	2.95	3.90	10.58	.06	.030
95%	1.96	1.98	2.01	2.02	2.04	2.06	2.09	2.13	2.18	2.23	2.26	2.31	2.36	2.45	2.57	2.78	3.18	4.30	12.71	.05	.025
96%	2.05	2.08	2.11	2.12	2.15	2.17	2.20	2.25	2.30	2.36	2.40	2.45	2.52	2.61	2.76	3.00	3.48	4.85	15.89	.04	.020
97%	2.17	2.20	2.23	2.25	2.28	2.30	2.34	2.40	2.46	2.53	2.57	2.63	2.71	2.83	3.00	3.30	3.90	5.64	21.21	.03	.015
98%	2.33	2.36	2.40	2.42	2.46	2.49	2.53	2.60	2.68	2.76	2.82	2.90	3.00	3.14	3.36	3.75	4.54	6.96	31.82	.02	.010
99%	2.58	2.63	2.68	2.70	2.75	2.79	2.85	2.95	3.05	3.17	3.25	3.36	3.50	3.71	4.03	4.60	5.84	9.92	63.66	.01	.005
99.9%	3.29	3.39	3.50	3.55	3.65	3.73	3.85	4.07	4.32	4.59	4.78	5.04	5.41	5.96	6.87	8.61	12.92	31.60	636.6	.001	.0005

Note. Round your *df* to the nearest tabled value. For *t* with *df*>>100, use *z*.

Adapted from Dawson, R.J.M. (1997). "Turning the tables: A t table for today." *Journal of Statistics Education* v.5, n.2.

Generated using Microsoft Excel.

For example, a survey is to have a margin of error of 10 percent. To calculate the confidence, subtract 10 percent from 100 percent (100 – 10 = 90 percent). The 90 percent (in the confidence column on the right of the table above) is used to find the t_α to be used in the Excel spreadsheet to calculate the population size.

The degree of freedom is calculated by subtracting 1 from the sample size: $n-1$ where n = the sample size. For a sample size of 20, the degree of freedom is 20 – 1 = 19. The chart does not have 19, so use the closest degree of freedom, which is 20. To find the t_α for a sample of 20 at a confidence of 90 percent, find the 90% row in the first column, and find 20 on that row. The t_α is 1.72. If the survey has a confidence of 95 percent and a sample of 100, the t_α is 1.98.