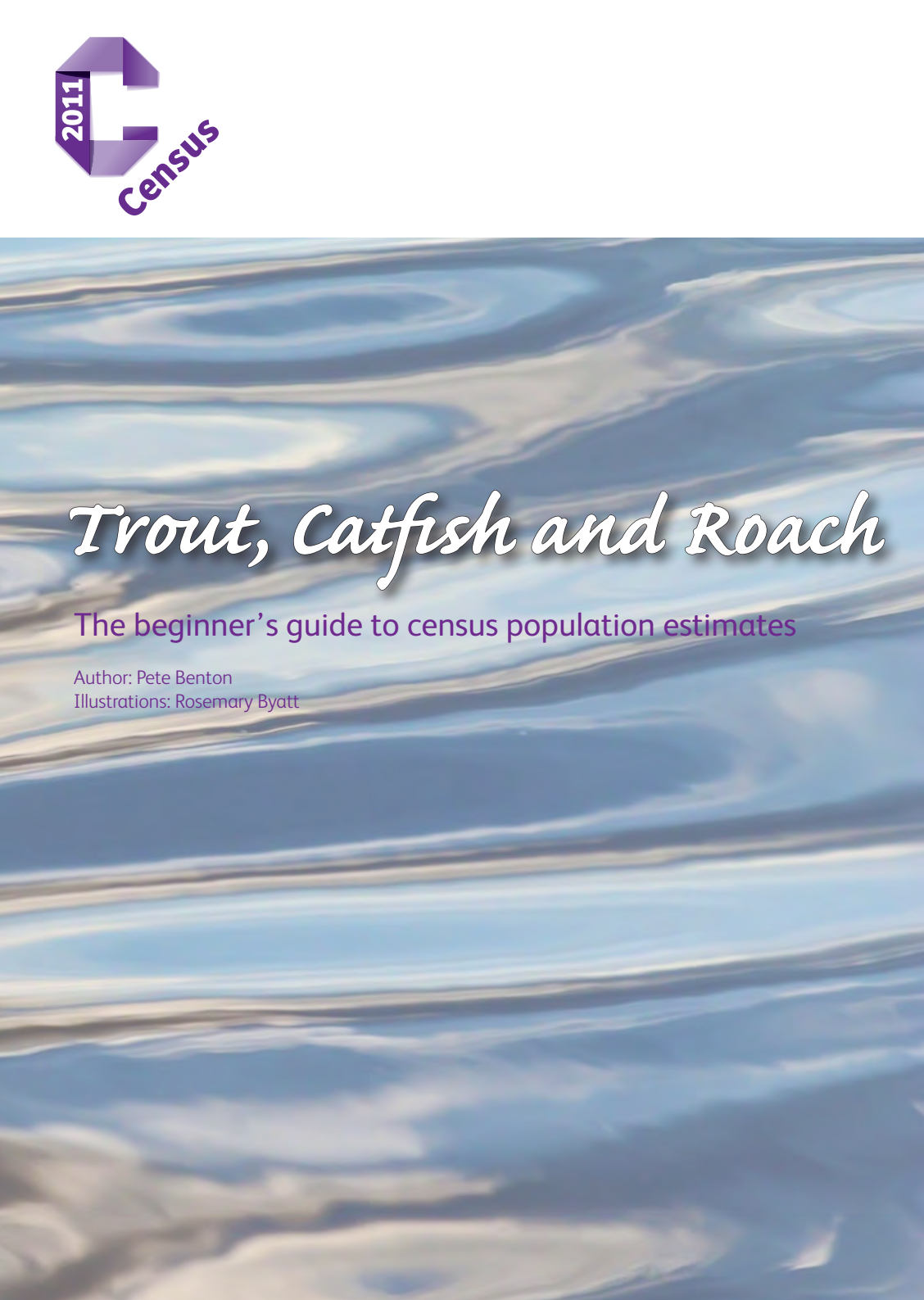


Trout, Catfish and Roach

The beginner's guide to census population estimates

Author: Pete Benton
Illustrations: Rosemary Byatt

Trout, Catfish and Roach ***is a beginner's guide to understanding population estimates and the role of the Census Coverage Survey. By using the analogy of fishing, this booklet takes readers through a series of easy equations to explain the way we reach our census population estimates. A worked example can be found at the end of this guide.***

Step one: counting trout

How do you find out how many trout there are in your pond?

You could drain it and count the fish I suppose, but it wouldn't do them much good. Perhaps if the pond was small you could try and catch them all. That would take a while. But there's nothing like fly fishing, so you decide to give it a go.

At the end of day one, you've caught 100 trout. Pretty good!

So how many are there in total? Still no idea, really. So you tag each one, put them all back (you've lovingly cared for them all of course), and carry on the next day.

The next day you manage to catch another 50; 25 with a tag, 25 without. So you've found another 25 and know for sure there are at least 125. Pretty good. That'll have to do.

But is that all you've found out?

Half of those you caught on day two already had a tag. Does that mean half the trout in the pond were tagged on day one? Roughly, yes.

And as you know you tagged 100 on day one, if half the trout in the pond were tagged, you can estimate that there must be around 200 in the pond altogether. *Give or take a few.*

The idea is simple when you take it step by step, but the implications are profound. You've only ever seen 125 fish in total, but can estimate with a fair degree of confidence that there are 75 or so more.

What does that have to do with the census?

We want to count everyone, but know we miss some. How many? Who knows? Well, we do actually.

We're not counting fish, we're counting people, but the principle is the same. We don't need to tag them - they all have names (and so that we can be sure, dates of birth and addresses too). We count them in the census, and then we do it again in what we call the Census Coverage Survey. This is called capture-recapture which, like the previous example, is used to estimate wildlife populations. It is also called *Dual System Estimation*.

We don't count as many second time of course, that would take too long (and our feet would get tired). But we know how many we counted first time and by matching the two lists of people we find out what proportion we counted twice. Just as we did with the trout, we can then estimate the total.

So is a census really as easy as counting fish? Not quite, of course. But that's enough to go on for now. To find out more, you can read on - and learn about catching roach, tench and catfish too.

Step two: counting roach, tench and catfish

So you know how to count trout. But you don't tend to catch roach, tench or catfish by fly fishing.

It would be tempting, if you didn't know better, to say there are only 200 fish in your pond. But as every fisherman knows, you have to know your fish. Trout eat flies. Roach prefer maggots.

So on day three, reinvigorated after an evening in the local pub, The Fisherman's Rest, you set out with a jar of maggots and try again. Lo and behold, there are other monsters of the deep to be found. Another 50 roach and tench, tagged and released as before. Very nice.

Day four yields another 20 roach and tench (The Fisherman's Rest beckons), only a quarter (five) with a tag this time. Having talked the idea through over a pint the night before, you're convinced you know how this thing works - there must be 200 roach and tench altogether. Plus the trout, that's 400. You're about right too.

But there's a nagging doubt. What about the catfish?

You know there's at least one because it scared your son witless at his birthday party the year before, but there's been no sign of it since.

Barmen are a fount of knowledge, and you learn that evening that the best time for catfish is at night - even better with a full moon. Maggots will do, just make sure they're still wriggling.

Spurred on by the amber nectar you head out and fish until the wee hours, bagging ten by sunrise. Hard work, but worth every minute. It crosses your mind to put one under your son's bed, but aged five it'll probably scar him for life. Instead you rest well and return in the evening of day five.

The next night yields ten more. But only one with a tag!

100 catfish, hiding, and no-one seemed to know they were there. Could it really be?

So, that's 500 fish altogether.

And what of the census?

In the census we don't have different types of fish, we have different types of people, some easier to *catch* than others. The vast majority of the population readily fill in their census questionnaire but others (like the catfish) need more careful attention - students and young men for example. And even with more

careful attention, like the catfish, we will have a lower *catch* rate than for the rest of the population. As long as we know which groups will be the most difficult, we can put in place additional measures to help us *catch* them - and we have, for example, very different processes in place to help improve response from students.

Catching as many as we can through targeted approaches for different population groups is clearly important, but even that isn't enough. Let's think some more about the fish.

Supposing after using our different techniques for catching the different types of fish, rather than estimating the total number of each type of fish separately, we just lumped them into one pot and did a single calculation. Would we get the right answer? Let's give it a go.

On day one we'd have tagged 160 fish (100 trout, 50 roach and 10 catfish). Day two would have seen a total of 80, 31 with a tag.

You'd need a calculator for this one, but 31 out of 80 had a tag, and we'd tagged 160 the day before, so you might estimate there to be 413 fish altogether. That's 160 x (80 *divided by* 31).

But we'd estimated 500 before!

What about the other 87? Where did it go wrong?

Well, we had different success rates for the different types of fish - a 50 per cent recapture rate for trout, 25 per cent for roach and tench, and only 10 per cent for catfish. Treating them all just as *fish* hides this fact. To get the right answer you have to add them separately, group by group.

In census terms we need to divide our calculation into groups (called *stratifying*). We know that response rates from young men are lower than middle aged men or older women, so we group our calculations by age and sex. Other factors drive response rates too, such as ethnic group, and there are other indicators we use, such as whether a person owns their home or rents it

(homeowners tend to be proactive responders). We use all these factors, and more.

So now you can count the fish in your pond. But what about your neighbour's pond?

Step three: beating your neighbour

Your neighbour is nice enough, but fishing in his pond wouldn't be on. However, you'd quite like to know if he has more fish than you.

One evening in The Fisherman's Rest you mention your love of fishing and he invites you and a friend round there and then. You can't believe your luck. Your neighbour really enjoys the beers you brought and, before you know it, it's dawn. 50 trout, 20 roach and tench, and three catfish. Not a bad haul. Sadly, he never invites you back. Tagging his fish perhaps wasn't the best of ideas.

Tantalised by the information you have gleaned, and never one to be beaten, you think through what you know. 50 trout, 20 roach and tench, and three catfish. And from your fishing experience in your own pond, you know trout are the easiest to catch (50 per cent recapture rate), roach and tench are harder (25 per cent) and catfish are hardest of all (10 per cent).

You haven't had the luxury of a second day's fishing next door but you can still guess that, having put in the same resources first time round (three people on the first day in both ponds), your initial capture rates should be about the same. And calculating separately for each type of fish, your recapture rates should stay the same too.

It takes you a while to get your head round the maths, but you estimate that he has 100 trout, 80 roach and tench, and 30 catfish: 210 fish in total.

You win! (Not that keeping up with the Jones's means anything to you)

Back to the census

Our second count doesn't happen everywhere. We only revisit 1 per cent of postcodes (or 1 per cent of ponds) but from the first day's fishing we've done in every single pond, we can use the capture/recapture rates (from the 1 per cent) to estimate the population of the 99 per cent.

But we're not quite there yet …

Step four: one for the really committed

Alright, this is pushing it a bit for a beginner's guide, but if you've read this far you're obviously keen and will tolerate a fisherman's tale …

Different types of pond

The fact is, your neighbour has quite a few ponds - and (you might have forgotten this fact) so do you.

Some are large, some are small. Some are deep, some are shallow. Some are stagnant and clogged with weeds whereas others, fed by streams, are crystal clear.

You think that trout love the clear ones and don't fare so well in the others, and you've heard that catfish love the weeds, but you're not quite sure.

Enticed by a good few more beers you invite some friends for a *day's fishing* (how generous they think you are). You choose one pond of each type and set

them off with flies and maggots, and easily persuade a few to stay and fish in the moonlight.

The day confirms your theory - different types of fish dominate the different types of pond. You hand out tags, the fish are released and you all head off to rest.

Persuading a few friends to fish the next day you find the ratio of tagged fish for each type of fish in each type of pond.

Just assuming your neighbour did the same (which, clearly, he wouldn't) and invited you and your friends for a day's fishing in his various types of pond, and assuming you classified his ponds into the same types as you'd previously classified yours, you could estimate his total fish population in each type of pond, and hence the total.

It would need a spreadsheet or two, not just a calculator, but you could do it.

And you remember the bit about *give or take a few*? Well, with only a few ponds in your garden to build from, it would be *give or take quite a lot*. You'd need a whole load of ponds in a whole load of gardens to build a good enough model to get back in the realms of *give or take a few* but if you did, then you could estimate for the rest of the town. You'd need to keep a good list of ponds and a good classification of pond types, loads of friends on day one, and quite a few on day two. The beer would be quite expensive, and I'm not sure that the spreadsheets would be up to it …

And what of the census?

Thankfully we have postcodes (ponds), interviewers (fishermen), statisticians and computing power.

The answers to the "what-ifs?" in the back of your mind

Once bitten twice shy: What if a fish, having being caught on day one, has learned about fishermen's hooks and stays away? Or in census terms, what if someone who has filled in their questionnaire refuses to take part in the Census Coverage Survey?

This is a real problem. If some of our tagged fish have learned their lesson, then (in our first example of 100 trout on the first day, 50 on the second, half tagged, half not), the proportion of untagged fish would be greater than if the tagged fish hadn't learned. We would overestimate the population. For this reason we choose to *use a net* second time, rather than use a rod and line. We do the coverage survey differently, using a face-to-face interview rather than a self-completion questionnaire.

Spotted catfish: What if there is a type of catfish that only eats worms? Flies and maggots won't help - we wouldn't catch them on either day.

This is true. There are two things we need to do:

- do our best to understand all the types of fish, and how to catch them;
- do our best to group (*stratify*) for the important types of fish.

Roach and tench were grouped together in our example - we assumed them to have the same capture / recapture rates. If they don't we'll get the wrong answer.

But how far do you go? Ultimately there are thousands of population sub-groups with slightly different capture / recapture rates. The *law of diminishing returns* applies. Dealing with the main sub-groups is generally sufficient.

Using a net on day two helps again. We might find there are some perch

around too, that don't like flies, maggots or worms (I'm not an expert fisherman - perch might well like maggots!).

We have done extensive research using 2001 Census and Census Coverage Survey data to understand the personal characteristics that underpin census non-response. This allows us to group and sub-group the 2011 Census data accordingly.

The Census Coverage Survey will also apply different methods. Two key differences are that it will use doorstep interviews rather than self-completion questionnaires and will not be conducted using address lists (field staff will simply be given an area boundary on a map and asked to find all households and people within it). This will also find addresses missed by the census address register development process and allow us to make adjustments for those too.

Ultimately things can go wrong, but in 2001 we had accurate results in over 95 per cent of local authorities. We have learned the important lessons from the 15 or so local authorities where things didn't work perfectly and have addressed the problems to improve practices in the most difficult areas. This will also help maintain, and possibly improve, overall quality in the other 95 per cent of areas.

Some ponds have bigger populations than others. Equal resources for each on day one doesn't necessarily result in the same capture rates on day one.

We don't actually control the number of census resources needed by postcode (or pond) - we control it by the number of addresses. Postcodes with more addresses will have more field staff, plus an adjustment for the level of difficulty of the area.

We will then monitor the response rate by area day by day during the census operation, and add more field staff to the areas with lower response rates, so that the first capture rate is similar everywhere (or at least, similar for all ponds of similar types).

Controlling this first capture rate is more effective in managing errors than allocating the same amount of resource to each postcode, as was described in the ponds example (a single fisherman per pond).

This is a significant jump forward from 2001 when we could only create areas for field staff based on estimated population size. In addition, through the questionnaire tracking system, we will now have real time response rates for each area. We can move field staff to ensure we get the best results from difficult areas.

How problematic is the *give or take a few*?

Clearly what we have is an estimate, not an exact count. Back to the first example, 100 trout on day one, 50 on day two, 25 with and 25 without a tag.

Supposing on day two we'd only caught two trout, one with a tag and one without. We'd still estimate that 50 per cent of the fish in the pond had a tag, and that there were 200 in total. But based on only two fish on the second day, it would be suspect.

Supposing we'd caught a third trout on day two, by chance. Depending on whether or not it had a tag, we'd either be estimating that we had 150 or 300 fish in total - quite a difference either way.

The key is in having a big enough sample second time round. What is big enough? Well, we estimate that we can say, with 95 per cent confidence, that our 1 per cent sample of postcodes gets us a census population estimate for England and Wales that is within 0.2 per cent of the true answer.

This is quite complex. Surely we can't rely on all these bits working perfectly?

Although we will have spent seven years designing these processes for the 2011 Census and are building on all the work prior to the 2001 Census, we still can't say that it will work perfectly. There is potential for errors to creep in. For this reason we have also put a huge amount of effort into quality assuring the results.

We use other sources of information such as pension, health, birth and school pupil records to give us further estimates of the size of the population's sub-groups; comparing our results with these other records can highlight potential areas of concern.

We also use demographic analysis, such as the ratio of men to women (sex ratios), mortality rates and fertility rates, each of which can flag up further issues to explore.

But what if there are some fish in the pond that cannot be caught with a rod or a net (for example, scallops in the mud at the bottom)? How do we estimate for them?

This is the really tricky bit for any census (or fishing) operation. There are some types of individuals and households who we will never count in a census (such as those who believe they have something to hide) nor will they take part in the Census Coverage Survey. Because they are never seen, they are not included within the estimates of the population using the capture/recapture methods described.

Some of these may not be covered by the other sources we use in quality assurance either.

If we believe that these types of individuals definitely live in *whole* households that are not counted, we might be able to use an external count of addresses (for example, the census address register or the council tax list) to make an adjustment. This is the approach we used in 2001.

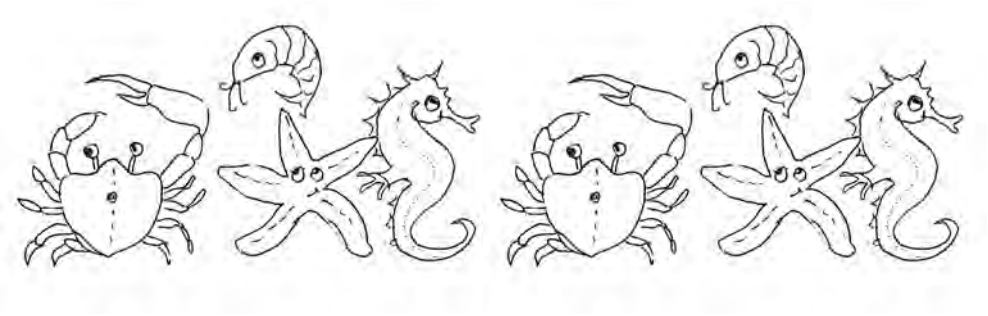
For a small random sample of records, we can also check back to the 2001 Census, and see whether we have missed people who were included then.

But this may well leave a final, small group who aren't included on any lists or in any data sources. This is where demographic analysis really comes into play.

In the UK we have excellent birth and death records which allow us to calculate reliable key demographic measures such as sex ratios, mortality rates and fertility rates. These can give us very important insights. Emigration and immigration complicate the sex ratio analysis so, ahead of the 2011 Census, we are researching sex ratio patterns of the UK-born population and migrant

flows from administrative and survey data sources. This allows us to develop a better idea of what we expect the sex ratio to be for each age group.

In some areas of the country we will work even more closely with local authorities to understand their local data. We can then tailor our census field work and field staff appropriately and use local insights to complement the quality assurance process.



Where can I find out more?

Further detail, including a more statistical description of the method, can be found at:

www.ons.gov.uk/census/2011-census.

(From the menu on the left click on consultations; user and advisory groups; census advisory groups; and statistical development. You can then download the 2011 UK Census Coverage Assessment details as a pdf.)

Further copies of *Trout, Catfish and Roach* are available by e-mailing census.communications@ons.gsi.gov.uk

Visit http://comms.census.gov.uk for general census information.

Step one: counting trout

Day one catch:

100 **trout (tagged and returned)**

Day two catch:

25 (already tagged) **+**

25 (newly caught)

25 (already tagged) **+**

25 (newly caught)

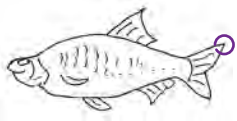
25 (already tagged)

$$\frac{25 + 25}{25} = 2$$

(first trout catch) **100 x 2** (ratio equation) = **200** trout estimated

Step two: counting roach, tench and catfish

Day three catch:



50

roach/tench (tagged and returned)

Day four catch:

5 (already tagged)

+

15 (newly caught)

5 (already tagged)

+

15 (newly caught)

5 (already tagged)

$$\frac{5 + 15}{5} = 4$$

(first roach/tench catch) **50 x 4** (ratio equation) = **200** roach/tench estimated

Total of fish estimated:

200

+

200

= 400

Day five catch:

10 catfish (tagged and returned)

.....

Day six catch:

1 (already tagged) + 9 (newly caught)

.....

1 (already tagged) + 9 (newly caught)

1 (already tagged)

$$\frac{1 + 9}{1} = 10$$

(first catfish catch) **10 x 10** (ratio equation) **= 100** catfish estimated

.....

Total of fish estimated:

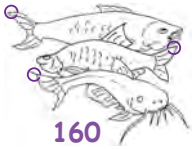
200 +

200 +

100 = 500

All fish into one pot

Day one catch:



160 **fish (tagged and returned)**

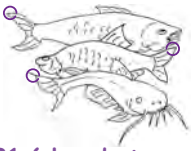
Day two catch:

31 (already tagged) +

49 (newly caught)

31 (already tagged) +

49 (newly caught)



31 (already tagged)

$$\frac{31 + 49}{31} = 2.58$$

(first fish catch) **160 x 2.58** (ratio equation) = **413** fish estimated