Capture-Recapture in the 2012 Census of Agriculture:

A Beginner's Guide

Using the analogy of fishing, this beginner's guide takes readers through a series of easy equations to explain the way we reach our census of agriculture estimates. A worked example can be found at the end of this guide.

Step one: counting bass

How do you find out how many bass 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 fishing with lures, so you decide to give it a go. At the end of day one, you've caught 100 bass. 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 bass in the pond were tagged on day one? Roughly, yes. And as you know you tagged 100 on day

one, if half the bass 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 of agriculture?

We want to count every farm, but know we miss some. How many? Who knows? Well, we do actually. We're not counting fish, we're counting farms, but the principle is the same. We don't need to tag them - they all have operators (and so that we can be sure, addresses too). We count them in the census of agriculture, but we also count them each year in the June Area Survey (JAS). 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 in the JAS of course, that would take too long (and our feet would get tired). But we know how many we counted in the census of agriculture and, by matching the two lists of agricultural operations, we find out what proportion we counted twice. Just as we did with the bass, 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 blue gill and catfish too.

Step two: counting blue gill and catfish

So you know how to count bass. But you don't tend to catch blue gill or catfish with lures.

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. Bass can be caught with lures. Blue gill prefer worms.

So on day three, reinvigorated after an evening at the local diner, The Fisherman's Rest, you set out with a can of worms and try again. Lo and behold, there are other monsters of the deep to be found. Turns out green sunfish also like worms. Another 50 blue gill and green sunfish, tagged and released as before. Very nice. Day four yields another 20 blue gill and green sunfish (The Fisherman's Rest beckons), only a quarter (five) with a tag this time. Having thought the idea through the night before, you're convinced you know how this thing works - there must be 200 blue gill and green sunfish altogether. Plus the bass, 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. The owner of The Fisherman's Rest is a fountain of knowledge, and you learn that evening that the best time for catfish is at night - even better with a full moon. Worms may be okay, but liver is better.

Spurred on by this new information, 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, at age 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 of agriculture?

In the census, we don't have different types of fish; we have different types of farms and farm operators, some easier to *catch* than others. The majority of the population of farm operators readily fill in their census of agriculture questionnaires, but others (like the catfish) need more careful attention – small farm owners and young farm owners 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 Native American operators living on reservations.

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 bass, 50 blue gill and green sunfish, 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).

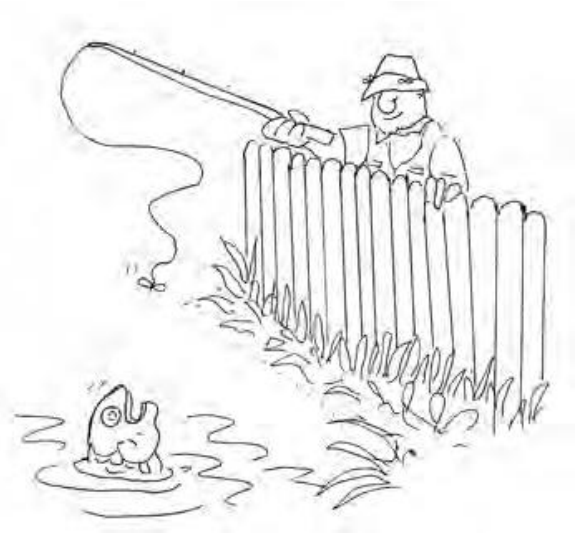
But we'd estimated 500 before! What about the other 87? Where did it go wrong?

Well, we had different success rates for the different types of fish - a 50 percent recapture rate for bass, 25 percent for blue gill and green sunfish, and only 10 percent 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 of agriculture terms, we need to divide our calculation into groups (called *stratifying*). We know that response rates from owners of small vegetable farms are lower than medium-sized corn farms or large cattle operations, so we group our calculations by size and type of farm. Other factors drive response rates too, such as the

age, sex, and race of the operator. We use all these factors, and more. So now you can count the fish in your pond. But what about your neighbor's pond?

Step three: beating your neighbor

Your neighbor is nice enough, but fishing in his pond wouldn't go over. 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 to come over there and then. You can't believe your luck. Your neighbor really enjoys the food and drink you brought and, before you know it, it's dawn. 50 bass, 20 blue gill and green sunfish, and three catfish. Not a bad haul. Sadly, he never invites you back. Tagging his fish perhaps wasn't the best of ideas.

Tantalized by the information you have gleaned, and never one to be beaten, you think through what you know. 50 bass, 20 blue gill and green sunfish, and three catfish. And from your fishing experience in your own pond, you know bass are the easiest to catch (50 percent recapture rate), blue gill

and green sunfish are harder (25 percent) and catfish are hardest of all (10 percent). 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, 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 math, but you estimate that he has 100 bass, 80 blue gill and green sunfish, 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 of agriculture

The JAS doesn't happen everywhere. We visit less than 1 percent (actually 0.6%) of the land area (or 0.6 percent 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 less than one percent) to estimate the population of the 99.4 percent. 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 neighbor 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 bass 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 good food and drink 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 lures, worms and liver, 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 neighbor 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 on your land 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 areas 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 food and drink would be quite expensive, and I'm not sure that the spreadsheets would be up to it …

And what of the census of agriculture?

Like ponds, the land areas vary across the U.S. Some are urban, some are a mix of urban and agricultural, and some are primarily agricultural. Corn farms tend to be in areas that are predominantly agriculture. Small farms may be more common in urban areas than in areas dominated by large agricultural operations. Just as bass are more common in some types of ponds than in others, some types of farms are more common in some areas than in others. As an example, you would be more likely to find a maple syrup farm in Vermont than in Arizona. Christmas tree farms are more likely to be found in North Carolina than in Nebraska, though they are in both states. Accounting for differences in the capture rates for different fish improves the estimates; similarly, accounting for differences in land areas leads to more precise census estimates.

Thankfully we have land areas (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 of agriculture terms, what if someone who has filled in their JAS questionnaire refuses to take part in the census of agriculture?

This is a real problem. If some of our tagged fish have learned their lesson, then (in our first example of 100 bass 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* the second time, rather than use a rod and line. We do the JAS and census of agriculture differently, using a face-to-face interview for the JAS and a self-completion questionnaire for the census of agriculture.

Spotted catfish: What if there is a type of catfish that only eats maggots? Lures, worms and liver 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.
Blue gill and green sunfish 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 subgroups 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 blue gill and green sunfish around too, that don't like lures, worms, liver or maggots (I'm not an expert fisherman – blue gill and green sunfish might well like maggots!).

We have done extensive research using 2007 Census of Agriculture and JAS data to understand the personal characteristics that underpin census nonresponse. This allows us to group and subgroup the 2012 Census of Agriculture data accordingly.

The JAS also applies different methods. Two key differences are that it uses in-person interviews rather than self-completion questionnaires and is not conducted using address lists (field staff are simply given an area boundary on a map and asked to find all agricultural operations within it). This will also find operations missed by the census mailing list development process and allow

us to make adjustments for those too. Ultimately things can go wrong, but in 2007 we had accurate results for most types of farms. We have learned the important lessons from the types of operations 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 for all types of farms.

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 county (or pond) - we control it by the number of operations. Counties with more agricultural operations will require more staff time, 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 of agriculture 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 county, as was described in the ponds example (a single fisherman per pond).

Is it easy to figure out whether or not an operation is a farm?

No – sometimes the JAS identifies an operation as a farm and the census of agriculture says it is not a farm, and vice versa. An operation must sell, or have the potential to sell, $1000 in agricultural products to be a farm. For small operations, it may not be clear whether or not an operation is a farm. So, if one person, say the husband, fills out the form, it may appear that the operation is a farm but, if the

wife fills out the form, it is found to not be a farm. Efforts are made to figure out whether or not an operation is a farm, but some are still not resolved. We have to account for these differences.

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 bass on day one, 50 on day two, 25 with and 25 without a tag. Supposing on day two we'd only caught two bass, one with a tag and one without. We'd still estimate that 50 percent 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 bass 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 the second time round. For this reason, the JAS sample is larger during census years.

How close are the estimates to the real numbers?

That depends on what is being estimated. Because bass have a higher recapture rate than catfish, we expect our estimates of the number of bass to be closer to the real number than our guess of the number of catfish. Standard errors provide a measure of how close the estimates are to what they are estimating. We are 95% confident that our estimates are within two standard errors. So, if our estimate of the number of farms in a county is 12 with a standard error of 2, we are 95% confident that there are between 8 (12 – 2 x 2) and 16 (12 + 2 x 2) farms in the county.

To reduce the number of standard errors published, a generalized coefficient of variation (generalized CV) is

reported for county-level estimates. The CV is the standard error divided by the estimate and, for the census of agriculture, a generalized CV is a CV that applies, at least approximately, for all counties in a state. So, if the generalized CV is 0.1 and the county estimate of the number of farms is 10, then the standard error is 0.1 x 10 = 1. In this case, we are 95% confident that the true number of farms in that county is between 8 (10 – 2 x 1) and 12 (10 + 2 x 1).

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

Although we will have spent five years designing these processes for the 2012 Census of Agriculture and are building on all the work prior to the 2007 Census of Agriculture as well as work by the U.S. Census Bureau, 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.

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 them?

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

Acknowledgments

This guide is adapted from *Trout, Catfish and Roach: a beginner's guide to census population estimates*., written by Pete Benton and illustrated by Rosemary Byatt (http://www.significancemagazine.org/details/webexclusive/1235233/Trout-Catfish-and-Roach-and-how-to-count-fish-and-non-fish.html) We are grateful to Owen Abbott for calling it to our attention and to the Office for National Statistics, UK, for allowing us to freely adapt it to meet our needs.

Step one: counting trout

Day one Catch:  **bass (tagged and returned)**

100

……

Day two Catch:  **+** 

25 (already tagged) + 25 (newly caught)

……

 **+** 

25 (already tagged) + 25 (newly tagged)
25 (already tagged)

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

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

Step two: counting blue gill, green sunfish and catfish

Day three Catch:  **blue gill/green sunfish (tagged and returned)**

50

……..

Day two Catch:  **+** 

5 (already tagged) + 15 (newly caught)

……..

 **+** 

5 (already tagged) + 15 (newly tagged)
5 (already tagged)

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

(first blue gill/green sunfish catch) **50 x 4** (ratio equation)
= 200 blue gill/green sunfish estimated

……..

Total of fish Estimated:  **+** 

200 + 200 = 400

Day five Catch:  **catfish (tagged and returned)**

10

.....

Day six Catch:  **+** 

1 (already tagged) + 9 (newly caught)

.....

 **+** 

1 (already tagged) + 9 (newly tagged) $\frac{1+9}{1}=10$
1 (already tagged)

(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:  **fish (tagged and returned)**

160

.....

Day two Catch:  **+** 

31 (already tagged) + 49 (newly caught)

.....

 **+** 

$$\frac{\text{31 (already tagged) + 49 (newly tagged)}}{\text{31 (already tagged)}} = \frac{31+49}{31} = 2.58$$

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