The largest juveniles are snails with no brood and no penis recorded. I believe that some of these are cases where there is an experimenter effect.

Figure 101a. Size variation over populations by maturity classes. The boxplot is based on median (horizontal black like), 25th and 75th percentiles (lower and upper hinges), minimum and maximum (lower and upper whisker), outlying points (black dots).

Model 1a: All samples combined

Fitting a model where the probability of maturity changes with size. Individuals classified as "immature" were treated as juvenile and no distinction was made between males and females.

```
s_mat <- function(size, mat, mean, slope) {
    logit_p <- (size - mean) / slope
    p <- exp(logit_p) / (exp(logit_p)+1)
    minusll <- -sum(dbinom(mat, 1, p, log = TRUE))
    return(minusll)
}
```

p is the probability that a snail will be adult, given its size (on natural logarithm scale), $mean$ is the average size at which snails become adult and $slope$ determines how variable the size at maturity is. There are two parameters in total and they are assumed to be constant across generations and populations.

Table 101. Maximum likelihood estimates of the parameters. The model was fitted to all the samples combined (Model 1a).

Parameter	Estimate	SE	
mean	1.91	0.02	***
slope	0.28	0.02	***

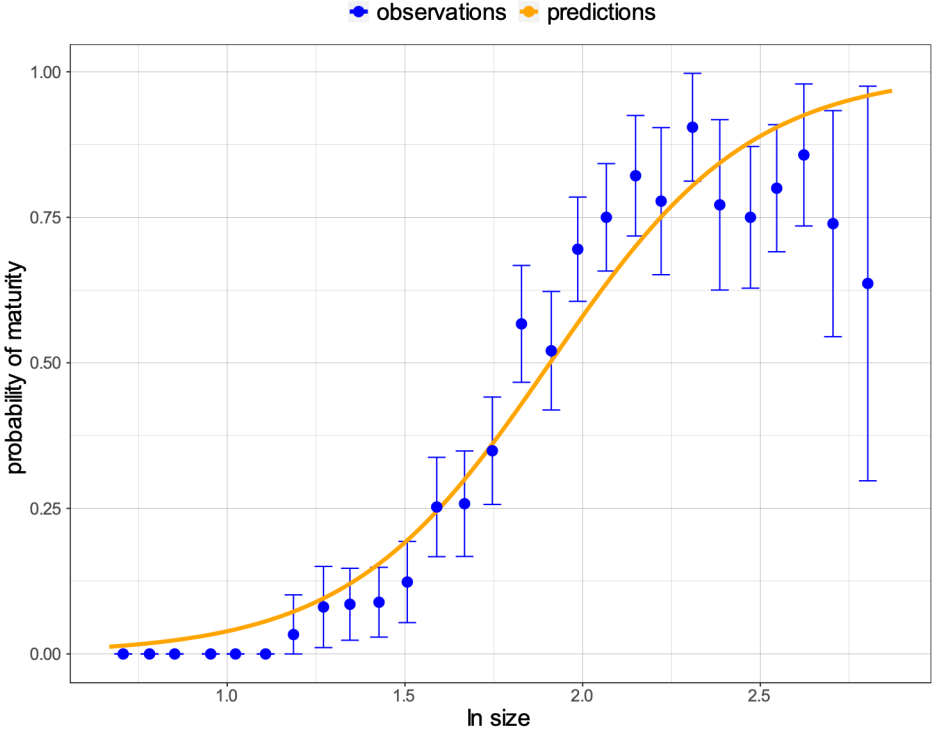Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 102a. Relationship between probability of maturity and size. Fitted curve in orange is superimposed on the observed proportions of mature snails(blue dots - proportions of adult snails for size bins. Blue error bars – 2.5th and 97.5th percentiles).

SE of the fitted curve is missing because I was not sure how to compute it. I know that for a simple logit model
```
dat <- data.frame(x = runif(20), y = rbinom(20, 1, .5), z = runif(20))
o <- glm(y ~ x, data = dat)
```

I can calculate SE at a given x as:
```
C <- c(1, x)
std.er <- sqrt(t(C) %*% vcov(o) %*% C)
```

based on the relationship
```
yhat = b0 + x*b1
```

But in our case
```
yhat = logit(p) = (x – mean) / slope
```

Can I still use the same formula for computing SE?

Model 1a: All samples combined

It is odd that the probability of maturity falls off for the largest size classes. This could be 'experimenter error' or, for field samples at least, it could be due to parasitism (sterilized individuals grow bigger). There is probably a case for cleaning out some of the outliers.

Perhaps parasitism can also be considered as "experimenter error" because the data do not include field samples that were recorded as parasitized. However, filtering parasitized field samples does not exclude those that might have been parasitized but parasites were not reported in the dissection sheets.

The juvenile class contains several snails that are as large as adults (Fig. 101a) and I expect that if I remove those samples, the probability of maturity does not fall off (Fig. 101b, 102b; Table 102).
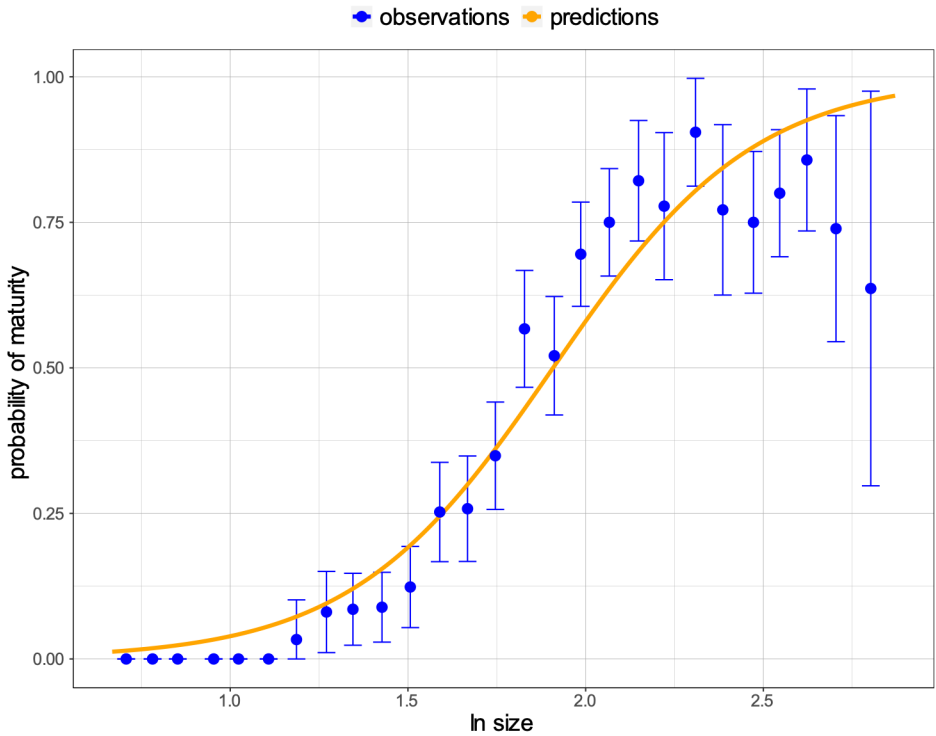

Figure 102a. Relationship between probability of maturity and size. Fitted curve in orange is superimposed on the observed proportions of mature snails(blue dots - proportions of adult snails for size bins. Blue error bars – 2.5th and 97.5th percentiles).

SE of the fitted curve is missing because I was not sure how to compute it. I know that for a simple logit model

```
dat <- data.frame(x = runif(20), y = rbinom(20, 1, .5), z = runif(20))
o <- glm(y ~ x, data = dat)
```

I can calculate SE at a given x as:
```
C <- c(1, x)
std.er <- sqrt(t(C) %*% vcov(o) %*% C)
```

based on the relationship
```
yhat = b0 + x*b1
```

But in our case
```
yhat = logit(p) = (x – mean) / slope
```

Can I still use the same formula for computing SE?

Table 101. Maximum likelihood estimates of the parameters. The model was fitted to all the samples combined (Model 1a).

Parameter	Estimate	SE	
mean	1.91	0.02	***
slope	0.28	0.02	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 1b: All samples combined without outliers

Mean size without removing outliers: 6.77 mm

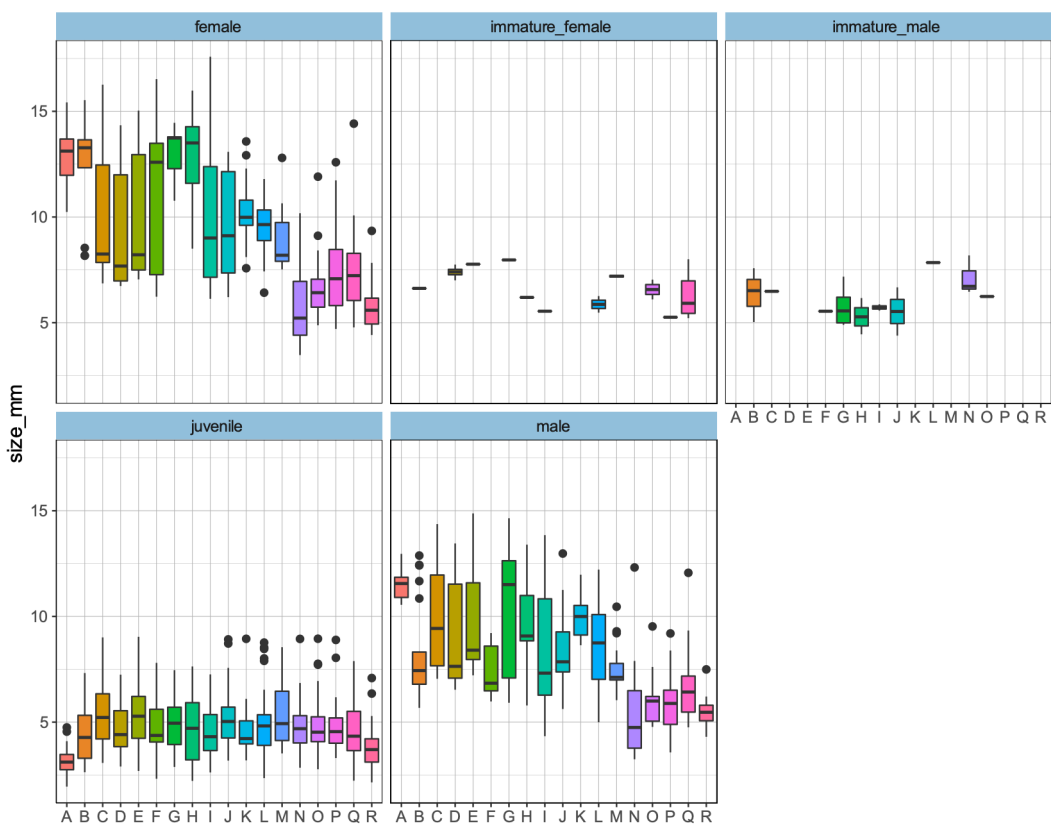Mean size if we remove 59 outliers: 6.52 mm

Figure 101b. Size variation over populations by maturity classes after outlier removal. The boxplot is based on median (horizontal black like), 25th and 75th percentiles (lower and upper hinges), minimum and maximum (lower and upper whisker), outlying points (black dots).
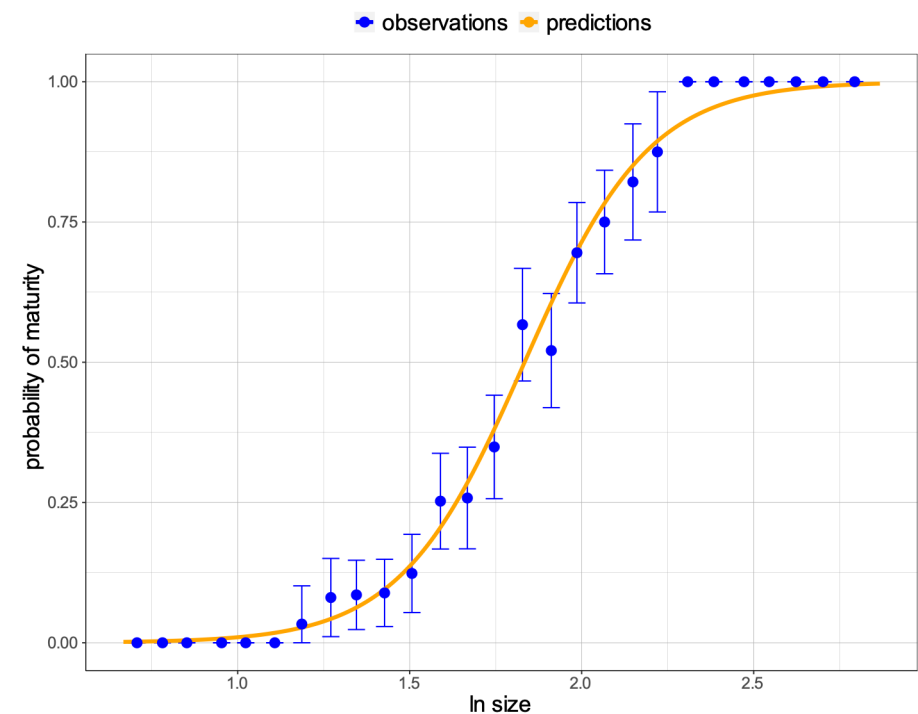

Figure 102b. Relationship between probability of maturity and size. Fitted curve in orange is superimposed on the observed proportions of mature snails(blue dots - proportions of adult snails for size bins. Blue error bars – 2.5th and 97.5th percentiles). Observed data did not contain "outlying" samples that were defined as falling beyond 1.5 × IQR (inter-quartile range or distance between the first and third quartiles).

Table 102. Maximum likelihood estimates of the parameters. The model was fitted to all the samples combined but without "outlying" samples (Model 1b).

Parameter	Estimate	SE	
mean	1.83	0.01	***
slope	0.18	0.01	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model two slopes and two sexes

Fitting a model where probability of maturity changes with size. Individuals classified as "immature" were treated as juvenile and adults were classified as either females or males.

What is the most sensible function
1) The same as Model 1 which can then be applied to two different data, one where males are removed and one where females are removed.
2) A function with two means and two slopes

```
s_mat <- function(size, mat, mean.f, slope.f, mean.m, s
  logit_pf <- (size - mean.f) / slope.f
  logit_pm <- (size - mean.m) / slope.m
  pf <- exp(logit_pf) / (exp(logit_pf)+1)
  pm <- exp(logit_pm) / (exp(logit_pm)+1)
```
But here I struggle because I do not know how to combine the *pf* and *pm* into *p*. This function should return the negative LL.
```
  minusll <- -sum(dbinom(mat, 1, p
  return(minusll)
}
```

Option 1 returns:

Parameter	Estimate		
mean.m		0.03	***
slope.m		0.02	***

Parameter	Estimate	SE	
mean.f	2.14	0.03	***
slope.f	0.29	0.02	***

I am also uncertain how to plot the observed data. My first try was to female + juvenile dataset and a male + juvenile dataset. Juvenile between the two datasets which means that the original individuals that were treated as juveniles are now included in both the juvenile dataset and the male + juvenile dataset. The same for immature males. **What are the alternatives?**
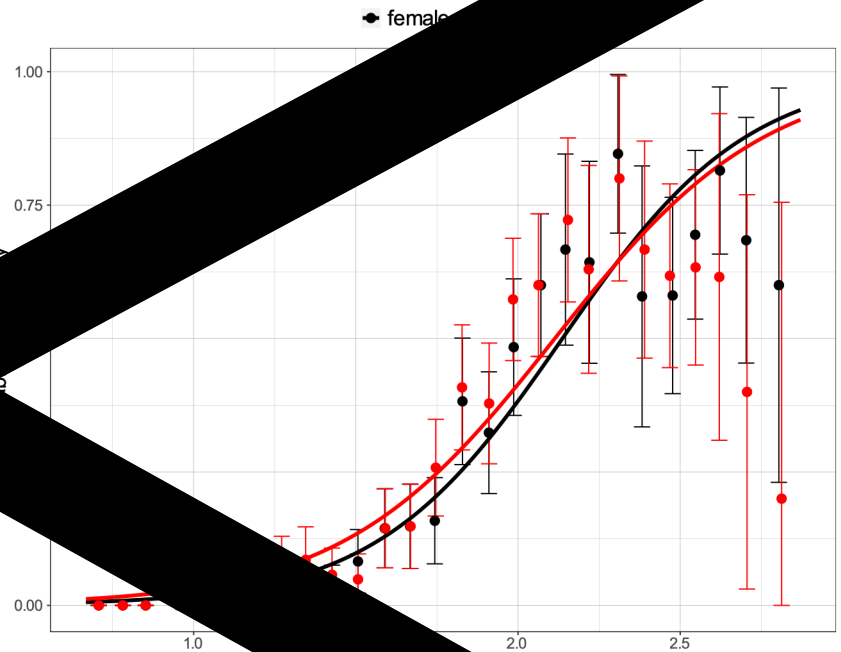

Figure 103. Relationship between probability of maturity and size in females and males. Fitted curve in black (females) and in red (males) superimposed on the observed proportions of mature females and mature males (black and red). Proportions of females and males for size bins, respectively. Error bars – 2.5th and 97.5th percentiles.

Actually, it looks like there might not be a sex difference, which would be rather surprising.

Model 2: four parameters and two sexes without outliers

Fitting a model where the probability of maturity changes with size. Individuals classified as "immature" were treated as juvenile and adults were classified as either females or males. The model is multinomial with three classes (female, male and juvenile) which means that, for example, the probability of being a mature male is the probability of being male and being mature. Hence, the addition of the sex ratio (0.5) in the calculation which can be replaced with the parameter sr (or $1 - sr$ for females). In order to keep sr between 0 and 1, the fitted parameter is logit sr. Other parameters are the average size at which males or females become adult (smat_m, smat_f) and a slope (b_m, b_f) that determines how variable the size at maturity is. We are assuming that the four parameters are constant across the contact zone.

```
s_mat_sex <- function(data, mean.f, slope.f, mean.m, slope.m) {

  logit_p_f <- (data[, "size_log"] - mean.f)  / slope.f
  logit_p_m <- (data[, "size_log"] - mean.m)  / slope.m

  p_f <- 0.5 * exp(logit_p_f) / (exp(logit_p_f)+1)
  p_m <- 0.5 * exp(logit_p_m) / (exp(logit_p_m)+1)
  p_j <- 1 - p_f - p_m
  # p_j <- (1 - p_f) * sr + (1-p_m) * (1 - sr)

  ll <- log(p_j)
  ll[data[, "maturity"] == 1] <- log(p_f[data[, "maturity"] == 1])
  ll[data[, "maturity"] == 2] <- log(p_m[data[, "maturity"] == 2])

  minusll <- -sum(ll)
  return(minusll)
}
```

Table 103. Count of samples for each maturity class. Outliers are excluded.

Juvenile (0)	Female (1)	Male (2)
771	315	328

Table 104. Maximum likelihood estimates of the parameters. The multinomial model was fitted to the data given three maturity classes (juvenile, female and male). Outliers were excluded.

Parameter	Estimate	SE	
mean female	1.87	0.02	***
slope female	0.17	0.01	***
mean male	1.80	0.02	***
slope male	0.19	0.02	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
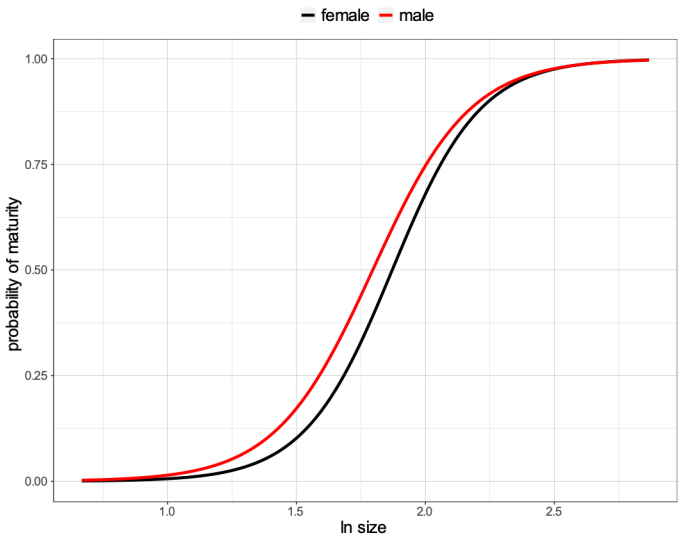

Figure 103. Fitted curves of the probability of being either female or male depending on size. Female in black and male in red.

Model fit comparison (AIC)

Between model 1b and model 2, the best fit to the data was obtained with model 1b (Table 105) suggesting that males and females in the experiment did not reach maturity at different sizes. Model 1a was not included here because it was fitted on a different dataset which contained outliers.

Table 105. Model fit comparison using AIC. Model 1b was the best fit and it included two parameters and two maturity classes (juvenile and adult).

Model	AIC
1b	1176.978
2	2065.903

After fitting this model, it could be extended to allow the parameters to vary over tanks (or with space in the clinal data). I did something a bit like this ages ago for ANG, but without separating males and females. It was possible to fit a cline for size at maturity even though we had rather few juveniles in that data set.

Model 3: four parameters and two generations without outliers

```r
s_mat_gen <- function(data, mean.p, slope.p, mean.o, slope.o) {

  logit_p_p <- (data[, "size_log"] - mean.p)  / slope.p
  logit_p_o <- (data[, "size_log"] - mean.o)  / slope.o

  p_p <- 0.5 * exp(logit_p_p) / (exp(logit_p_p)+1)
  p_o <- 0.5 * exp(logit_p_o) / (exp(logit_p_o)+1)
  p_j <- 1 - p_p - p_o
  # p_j <- (1 - p_f) * sr + (1-p_m) * (1 - sr)

  ll <- log(p_j)
  ll[data[, "maturity"] == 1] <- log(p_p[data[, "maturity"] == 1])
  ll[data[, "maturity"] == 2] <- log(p_o[data[, "maturity"] == 2])

  minusll <- -sum(ll)
  return(minusll)
}
```
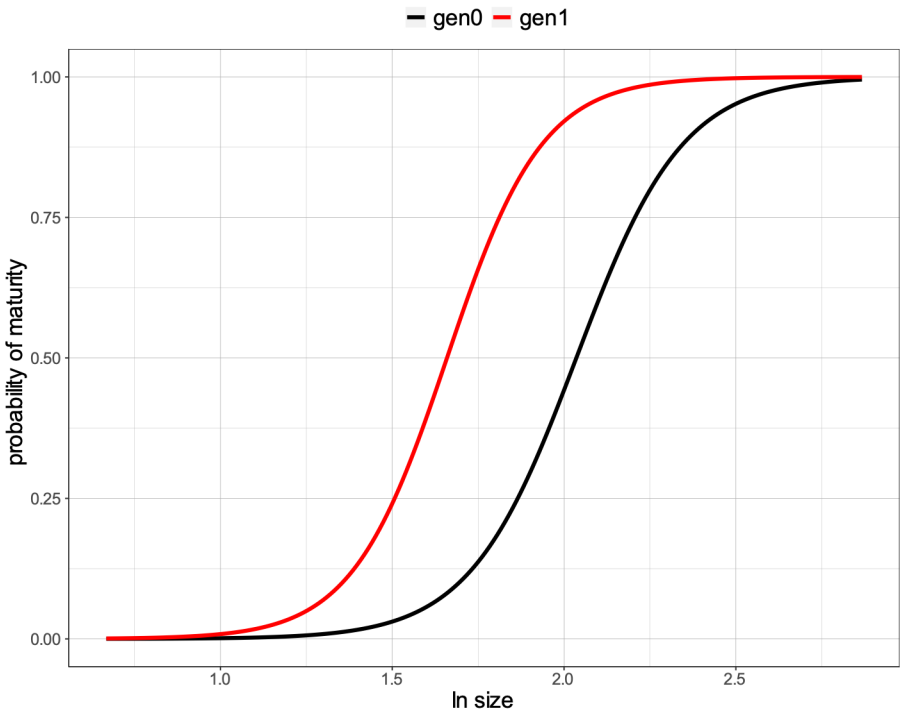
Table 106. Count of samples for each maturity class. Outliers are excluded.

Juvenile (0)	Parent (1)	Offspring (2)
771	274	369

Table 107. Maximum likelihood estimates of the parameters. The multinomial model was fitted to the data given three maturity classes (juvenile, adult parent and adult offspring). Outliers were excluded.

Parameter	Estimate	SE	
mean parent	2.04	0.02	***
slope parent	0.16	0.01	***
mean offspring	1.66	0.02	***
slope offspring	0.14	0.01	***

Figure 104. Fitted curves of the probability of being either a mature parent or a mature offspring depending on size. Parent in black and offspring in red.

Model fit comparison (AIC)

Between model 1b, model 2 and model 3, the best fit to the data was obtained with model 1b (Table 108) suggesting that the probability at maturity does not differ between sexes nor generations. Model 1a was not included here because it was fitted on a different dataset which contained outliers.

Table 108. Model fit comparison using AIC. Model 1b was the best fit and it included two parameters and two maturity classes (juvenile and adult).

Model	AIC
1b	1176.978
3	1950.473
2	2065.903

It was noticeable that in both model 2 and model 3, the most different parameter was *mean* and not *slope*. I assessed whether models with two *mean* parameters and one *slope* parameter improved the fit (Table 109). They do but not significantly (ΔAIC < 2).

Table 109. Model fit comparison using AIC. Model 1b was the best fit and it included two parameters and two maturity classes (juvenile and adult). The two models with "one slope" contained three parameters in total, two means and one slope.

Model	AIC
1b	1176.978
3 one slope	1949.326
3	1950.473
2 one slope	2064.434
2	2065.903

Checkpoint

1) Does the slope differ between sexes? NO, model 2 two slopes and model 2 one slope showed similar AIC values

2) Does the slope differ between generations? NO, model 3 two slopes and model 3 one slope showed similar AIC values

We can compare size at 50% maturity between generations and among tanks using a single slope for the logistic regression of maturity on size. We fitted a model for each population of generation 0 and generation 1 separately in order to predict size at 50% maturity and its standard error to use in the plasticity analysis.

However, this is not really possible!
When we removed the size outliers, most of the juveniles in generation 0 were excluded because too large compared to the median of the juvenile size distribution (Fig. 101a-b). The only populations of generation 0 with juveniles are J, L, M, N, O, P and Q and for these populations, size at 50% maturity was estimated with low standard error (highlighted rows in Table 110).

Can we instead use the fitted values from the best model (Model 1b) and take the average predicted size $\pm$ SE for each population in generation 0 and generation 1?

Table 110. Predicted size at maturity among populations of generation 0. The estimated parameter (Mean) and its standard error (SE) define the ln size at 50% maturity. The slope parameter was fixed to 0.18 (Model 1b).

Generation	Population	Mean	SE
0	A	-111.87	0.00
0	B	-1.63	5057.55
0	C	-1.55	3789.42
0	D	-1.53	3769.77
0	E	-1.47	3426.81
0	F	-1.70	5415.01
0	G	-1.69	4896.55
0	H	-1.76	4090.33
0	I	-1.70	4531.26
0	J	1.80	0.19
0	K	-1.73	3574.06
0	L	2.05	0.09
0	M	1.72	0.14
0	N	1.67	0.13
0	O	1.70	0.11
0	P	1.37	0.13
0	Q	1.57	0.11
0	R	-2.36	3814.38

Model 4: 18 parameters, one *slope* plus 17 *means*

Table 110. Count of samples for each maturity class. Juveniles were given maturity of zero and adults were given maturity from one to 18 depending on the population which they belonged to. Outliers are excluded ang both generations are combined.

Juvenile (0)	Adult A (1)	Adult B (2)	Adult C (3)	Adult D (4)	Adult E (5)	Adult F (6)	Adult G (7)	Adult H (8)	Adult I (9)	Adult J (10)	Adult K (11)	Adult L (12)	Adult M (13)	Adult N (14)	Adult O (15)	Adult P (16)	Adult Q (17)	Adult R (18)
771	15	34	39	37	35	36	23	22	37	40	46	28	41	42	39	49	34	46

single slope for the logistic regressions of maturity on size, in order to focus on comparing the size at 50% maturity among tanks. Probably this should be done separately for gen0 and gen1, where there might be biological reasons for slopes to differ, but perhaps that difference, and the difference between sexes, could first be tested on the full data set. Then the slope could be fixed and individual tank/gen combinations analyses separately to find the size at 50% maturity and its standard error to use in the plasticity analyses (or you could fit a model with one slope and means for different tanks).

Model 1: All samples combined

Given the logistic fit, a value for 'size at maturity' is the intercept (1.91 +/- 0.02 in slide 2) because this is the size at 50% probability of maturity (zero on the logit scale). If we had such an estimate for each tank and each generation, this could be treated as a trait in the plasticity analysis, just like weight, boldness etc. Adding tank to the analysis would give these separate estimates, I guess, or there might be enough data to do the analysis separately for each tank (and generation) (Model 3).
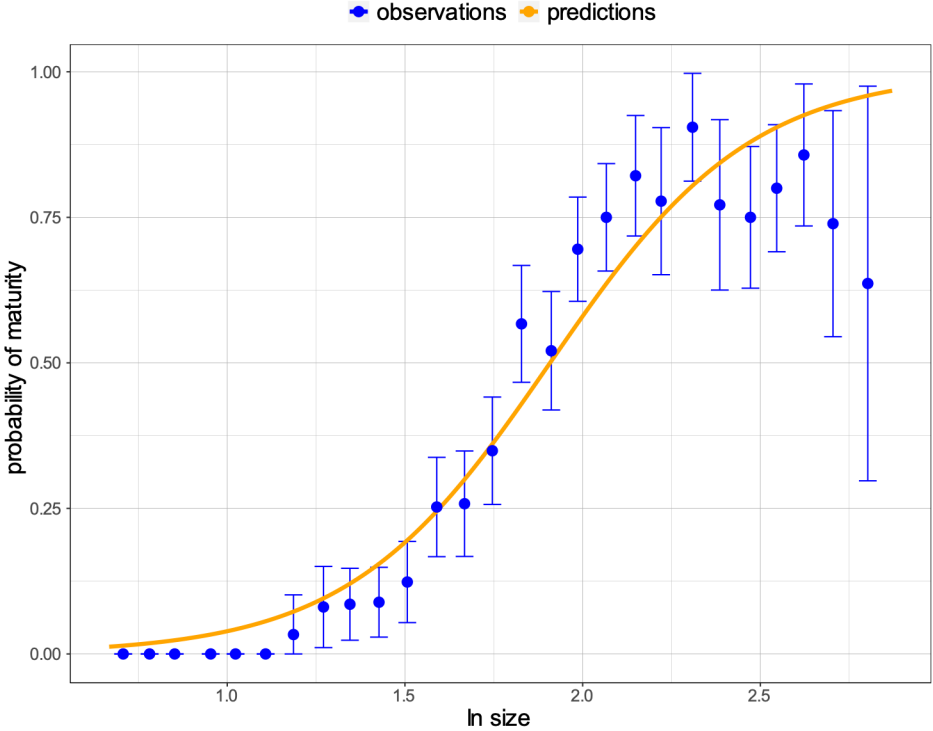

Figure 102a. Relationship between probability of maturity and size. Fitted curve in orange is superimposed on the observed proportions of mature snails(blue dots - proportions of adult snails for size bins. Blue error bars – 2.5th and 97.5th percentiles).

SE of the fitted curve is missing because I was not sure how to compute it. I know that for a simple logit model

```
dat <- data.frame(x = runif(20), y = rbinom(20, 1, .5), z = runif(20))
o <- glm(y ~ x, data = dat)
```

I can calculate SE at a given x as:

```
C <- c(1, x)
std.er <- sqrt(t(C) %*% vcov(o) %*% C)
```

based on the relationship

```
yhat = b0 + x*b1
```

But in our case

```
yhat = logit(p) = (x – mean) / slope
```

Can I still use the same formula for computing SE?

Table 101. Maximum likelihood estimates of the parameters. The model was fitted to all the samples combined (Model 1).

Parameter	Estimate	SE	
mean	1.91	0.02	***
slope	0.28	0.02	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1