



Estimating abundance

What's next

- We will extend the approach we developed for occupancy to abundance
 - Substitute surveys where we record presence-absence of a species with those where we record the number of individuals observed (counts)
- Evaluate whether a model fits the data



Perhaps the most common question asked of biologists:
How many are there?



Bev Doolittle

Abundance

Demographic parameter considered often to:

- Answer basic ecological questions
- Gauge the effectiveness of conservation and management strategies
- Monitor trends (e.g., estimate λ)

Complete counts of animals on sample units are rarely possible, so for inferences about abundance to be reliable, we address the potential biases introduced by imperfect surveys

Survey effectiveness or detection probability often varies in space and time, which introduces biases to survey data

Definitions

- **Abundance**: no. individuals in a population — used interchangeably with population size
 - Identical to the quantity **Population Total** (τ) in general sampling theory
 - Often denoted as N , so be careful not to confuse it with the no. sample units in the sampling frame (also N)
- **Density**: abundance **per unit area** (1.2 individuals/ha)



General approaches to estimate abundance

All Individuals Observed — Complete Counts ($p = 1$)

- Complete count or 'census' of the entire population
- Complete count on sample units (i.e., sampling)

All Individuals Not Observed — Incomplete Counts ($p < 1$)

- Count-based methods — no marking
 - Replicate Counts (= point counts)
 - Distance Sampling: Line and point transects
- Capture-based methods
 - Capture-Recapture
 - Removal Sampling

General framework to estimate abundance

Expected no. individuals counted on a survey, C , is related to true abundance, N , through p which can range from 0 – 1:

$$E(C) = p N,$$

To convert C into an estimate of abundance: $\hat{N} = \frac{C}{\hat{p}}$

Canonical
abundance
estimator



When surveys are imperfect, probability of observing an individual that is present, p , is < 1

Example: Count 20 quail on a plot; if $p = 0.5$, then:

$$\hat{N} = 20/0.5 = 40$$

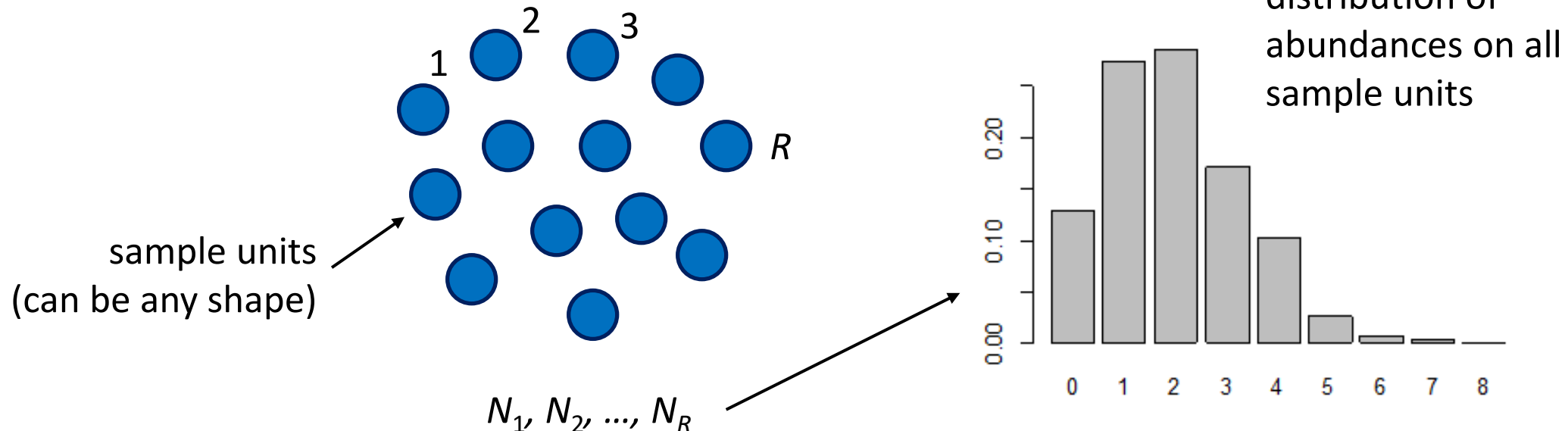
Why choose abundance over occupancy?

- Abundance is likely to change before distribution changes (i.e., before a local extinction)
- Abundance offers insight to [habitat quality](#); instead of 0, 1 we expect a range of values on sample units 0, 1, 2, 3, 4...
- In occupancy studies, [detection probability can be influenced by abundance](#), which can bias occupancy estimates
- Occupancy implies that $N > 0$, therefore occupancy can be estimated with abundance data

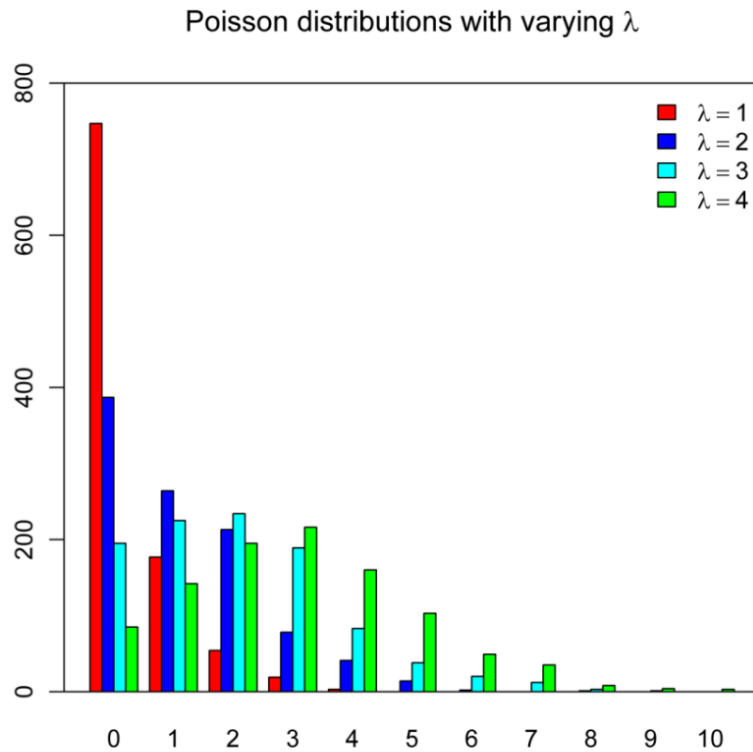
Abundance: single-season, replicate counts

Model based the idea of a 'meta-population', which represents the population of interest as a collection of smaller populations defined by the number of individuals, N , on each sample unit

Note: This is not a biological metapopulation



Poisson distribution



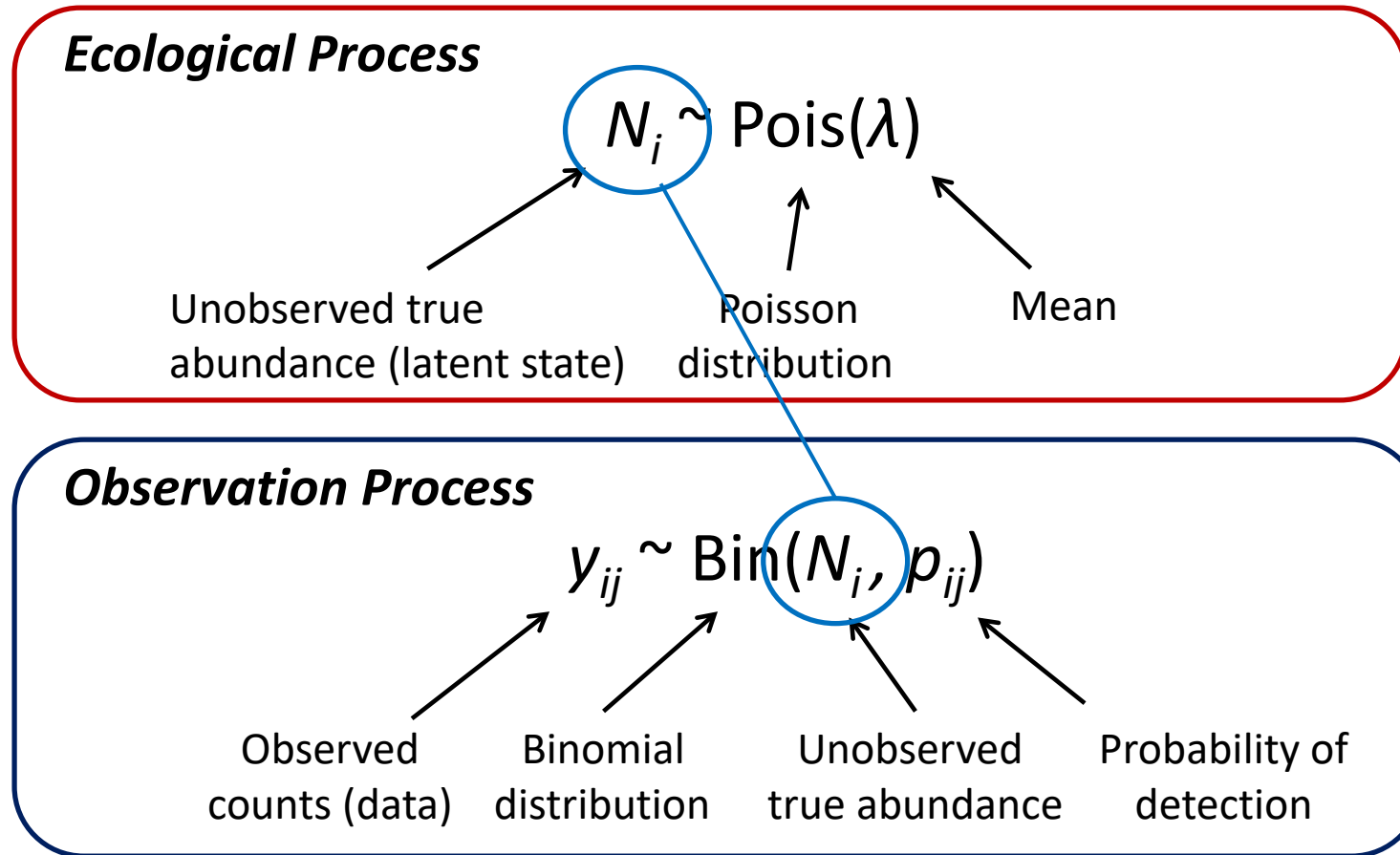
Discrete distribution used to represent expected counts of the number of events that occur within a fixed time period (i.e., a Poisson process)

Notation: $\text{Pois}(\lambda)$

One parameter:

$\lambda = \text{mean ("scale")} = \text{variance}$

Abundance: single-season, replicate counts



Incorporating covariates

GLMs account for sources of heterogeneity in **abundance** and **detection** processes

Model for the **abundance process (latent)**:

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$\text{link} = \log$$

$$\log(\lambda_i) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots$$

“Poisson regression”

Model for the **detection process** is similar to occupancy:

$$y_{ij} \sim \text{Binomial}(N_i, p_{ij})$$

$$\text{link} = \text{logit}$$

$$\text{logit}(p_{ij}) = \alpha + \beta_1 x_{ij} + \beta_2 x_{2ij}$$

“Logistic regression”

Combined, these are
the “*N*-mixture model”

Common distributions for abundance data

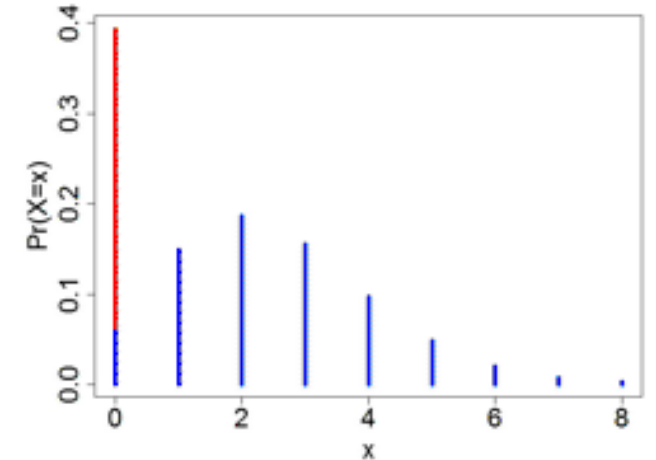
$N_i \sim \text{Poisson}$

$N_i \sim \text{Negative Binomial}$

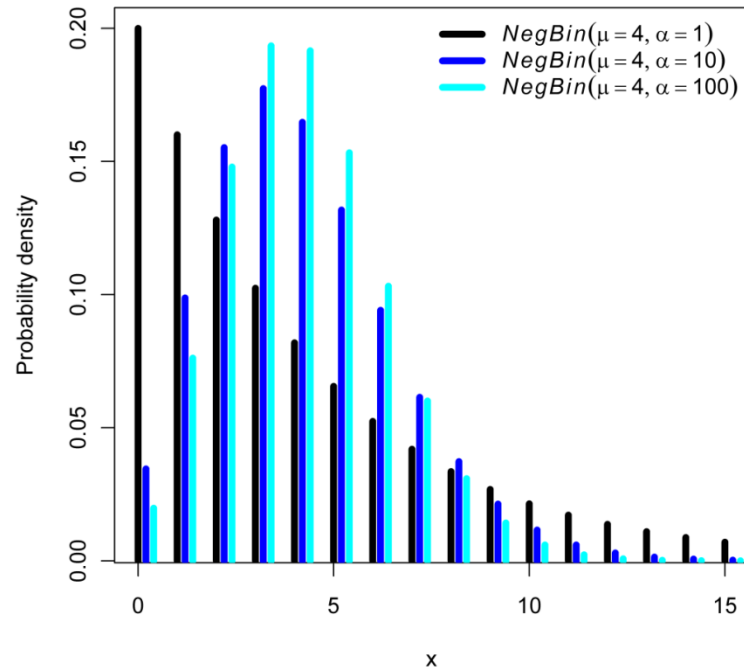
$N_i \sim \text{Zero-inflated Poisson}$

ZIP is a **mixture** of two distributions that represent two stochastic processes:

1. a Poisson distribution that generates counts, some of which may be zero
2. a binary distribution that generates excess zeroes



Negative binomial distribution



Discrete distribution used to represent expected no. trials necessary for x successes

Notation: $NegBin(\mu, \alpha)$

Two parameters:

μ = mean

α = dispersion

More flexible than Poisson because of variance (dispersion) parameter

When dispersion is low, NB is similar to Poisson

Example data: Swiss bird surveys



Willow tit data for 15 of 237 1 *km*² quadrats:

- counts --			--abun covs-----			---- sampling/observation covariates -----								
c.1	c.2	c.3	elev	forest	length	day.1	day.2	day.3	dur.1	dur.2	dur.3	int.1	int.2	int.3
1	0	0	1300	32	6.1	35	58	75	260	270	290	42.6	44.3	47.5
2	1	0	1270	66	4.5	28	39	61	176	145	150	39.1	32.2	33.3
0	0	0	380	45	6.2	21	40	86	180	160	165	29.0	25.8	26.6
3	3	1	550	31	6.9	20	42	63	195	240	270	28.3	34.8	39.1
0	0	0	390	8	4.6	35	50	75	150	130	140	32.6	28.3	30.4
0	1	0	1380	78	3.7	54	75	99	160	160	150	43.2	43.2	40.5
0	0	0	530	37	5.7	30	50	76	225	210	225	39.5	36.8	39.5
0	0	0	1190	18	4.0	26	47	71	165	175	240	41.2	43.8	60.0
1	1	1	1490	54	5.5	47	62	77	300	325	321	54.5	59.1	58.4
0	0	0	920	6	4.3	20	40	61	170	90	98	39.5	20.9	22.8
0	0	0	620	3	3.4	17	46	61	270	271	215	79.4	79.7	63.2
0	0	0	540	27	4.2	15	45	66	180	240	180	42.9	57.1	42.9
1	0	1	820	56	5.3	37	60	90	290	315	295	54.7	59.4	55.7
2	3	4	1220	66	5.6	45	61	86	225	240	255	40.2	42.9	45.5
1	1	2	1180	43	5.1	39	60	83	325	365	350	63.7	71.6	68.6
0	0	0	730	40	6.2	23	45	58	390	391	406	62.9	63.1	65.5

survey data

site covariates

for abundance

process

survey covariates for observation process

Example: elf owls

Point-counts surveys, monthly

4 Routes

18 plots/points per Route

Ecological Process

$$N_i \sim \text{Pois}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta_1 \text{Route}$$

Observation Process

$$y_{ij} \sim \text{Bin}(N_i, p_{ij})$$

$$\text{logit}(p_{ij}) = \alpha + \beta_1 \text{Week}$$

- Model selection
- Assess model fit
- Interpret parameter estimates



unmarked function for N -mixture models: **pcount ()**

Assess model fit

After selecting a model for inference, we must ask “does the model fit the observed data acceptably?”

Alternatively: “does it seem likely that the observed data arose from processes described by the model?”

Often use a **goodness-of-fit test** – two common approaches:

- Chi-square test – difference between **observed** values and those **expected** under the model
- Bootstrapping

Bootstrap

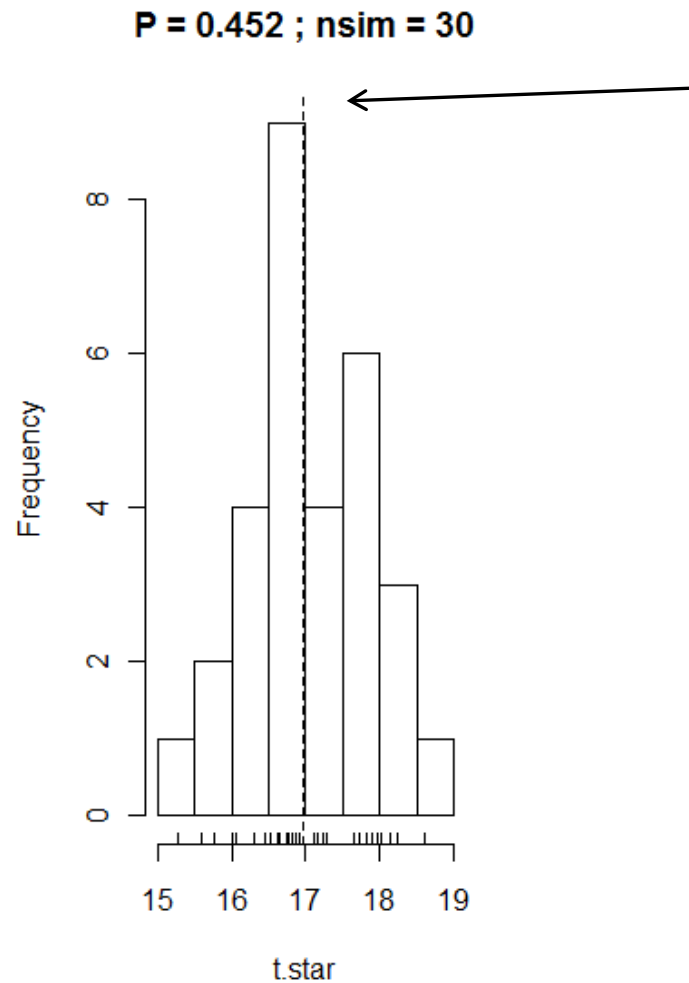
Evaluate fit by establishing the sampling distribution of an estimate based on data from a single sample

- Draw many “bootstrapped” samples from original data (= resampling)
- Fit the model to each sample and generate an estimate
- Create a sampling distribution of bootstrapped estimates
- Compare observed result to the bootstrapped sampling distribution

Allows us to gauge whether residual (error) variance in the observed data differs from the residual variance we'd expect if data arose under the model

unmarked function: `parboot ()`

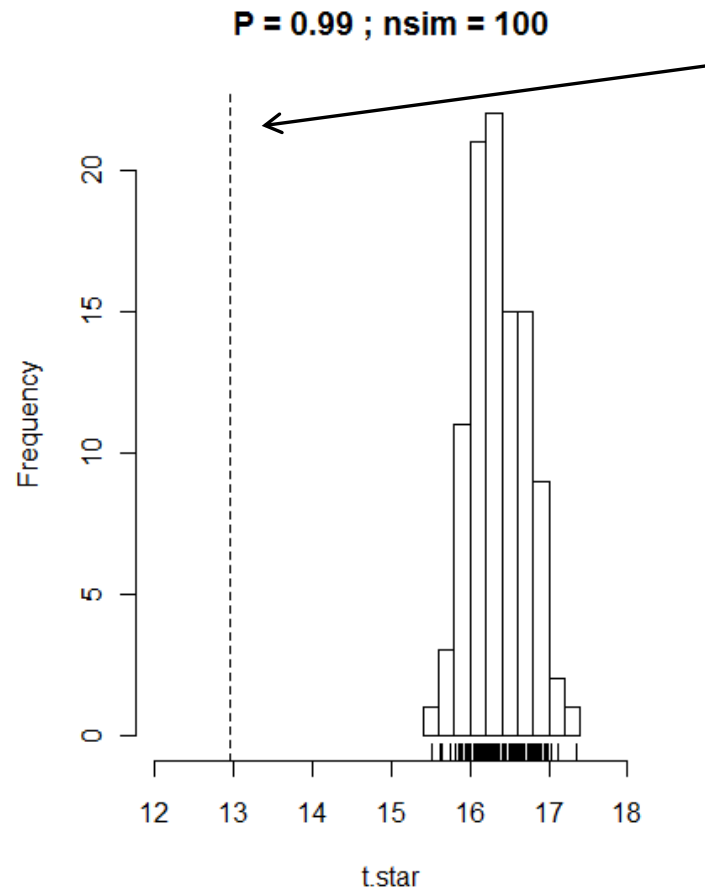
Assess model fit



Observed residual variance (dashed line) is similar to residual variance **expected** under the model

If $P > 0.10$ and < 0.90 , so conclude the model fits the data

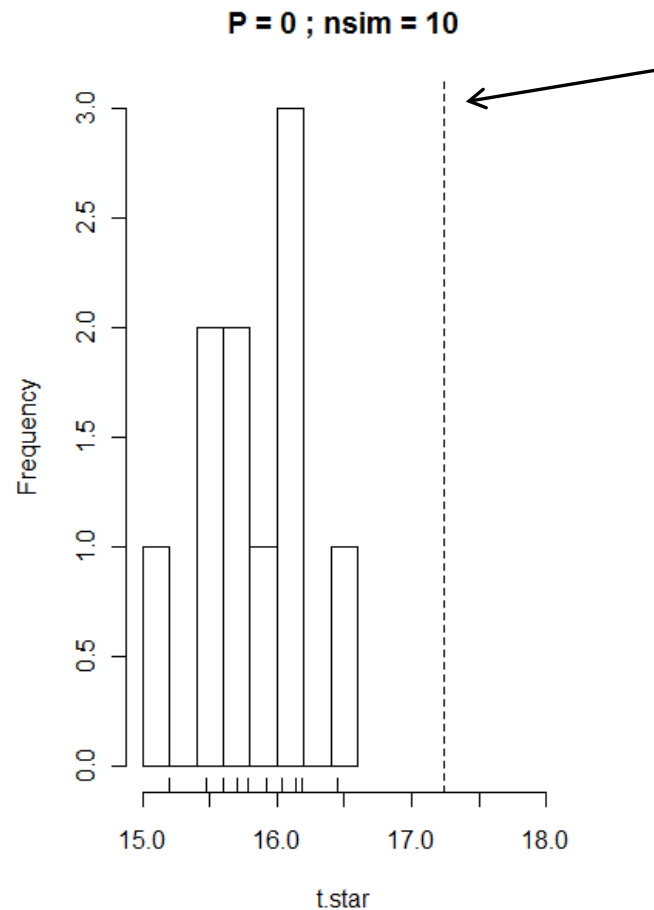
Assess model fit



Model is **underdispersed**; that is, **observed** variance is $<$ variance **expected** under the model

$P > 0.90$, so conclude the model does not fit the data

Assess model fit



This model is **overdispersed**; that is, **observed** variance is $>$ variance **expected** under the model

$P < 0.10$, so conclude the model does not fit the data

If the model does not fit, we can try:

- models with additional covariates
- alternative distributions for abundance, such as the Negative Binomial