A Bayesian approach to produce 100 m gridded population estimates using census microdata and recent building footprints

WorldPop, University of Southampton

17 November 2020

Contents

1 Introduction

Gridded population estimates can provide an important resource for planning government services, housing and population censuses, health and education initiatives, household surveys and other programes in situations when their are no recent census data available. Gridded population estimates can be aggregated to estimate total populations for custom areas suited to individual project goals. WorldPop uses two general approaches for producing gridded population estimates: bottom-up and top-down (Wardrop et al. 2018).

Top-down methods (Stevens et al. 2015) dissaggregate known population totals for administrative units into gridded population estimates at higher spatial resolution (e.g. 100 m). This approach is best for situations when reliable population totals for administrative units are available from a recent census or census projections. The population totals are dissaggregated based on relationships with gridded geospatial covariates that are mapped consistently across an entire study area or region of interest such as nighttime lights, impervious surfaces, and distance to city centers. Top-down approaches are not reliable when population totals from a recent census are not available and population projections are not reliable due to population displacement, time since the last census, or other reasons.

Bottom-up methods (Leasure et al. 2020b) use population counts from a sample of locations (e.g. from household surveys) to produce gridded population estimates with full national coverage. This approach is best for situations when population totals from census are not available but recent household survey data are available with full enumeration of people in a sample of clearly defined survey locations. Like the top-down method, populations are estimated based on relationships with gridded geospatial covariates. Bottom-up methods are often based on Bayesian statistical models that provide robust estimates of uncertainty, whereas top-down models based on machine learning models (Stevens et al. 2015) do not provide any confidence bounds on population estimates. One challenge with the bottom-up approach is that it requires geolocated household survey data that can be difficult to access due to privacy safeguards that must be put in place.

A method is needed that can produce robust gridded population estimates in situations when recent census data and geolocated household survey data are both unavailable. There are publicly available anonymized household survey data without specific location information that could be used for this purpose (e.g. IPUMS, DHS). Increasing availability of building footprints mapped from satellite imagery also provide a valuable source of information. A statistical method that measures statistical uncertainty for population estimates would be needed because the inherent uncertainty in population data that are not geolocated will likely produce population estimates with wide confidence intervals that would be important for end users to be aware of.

Our goal here was to develop a new Bayesian statistical model to estimate the number of households per building and people per household at sub-national spatial scales. We prioritized the use of publicly available data so that the method could be replicated across countries more easily. We demonstrated this new approach with a case study to produce gridded population estimates for Ghana. Our aim was to produce:

1. 100 meter gridded population estimates with national coverage,
2. Estimates of age-sex structure by region and settlement type,
3. Estimates of people per household by geographic unit and settlement type,
4. Estimates of households per building by geographic unit and settlement type, and
5. Robust estimates of uncertainty for population estimates (and other parameters) at any spatial scale.

2 Methods

The total population for an area can be described as a function of the number of buildings:

$$population = buildings \times \frac{households}{building} \times \frac{people}{household}$$

We develop this equation into a hierarchical statistical model below, but first we will describe the sources of information needed to inform the parameters on the right side of the equation.

2.1 Data

Building footprints. We used rasterized counts of buildings within approximately 100 m grid cells across Ghana (Dooley et al. 2020). These were derived from building footprints for sub-saharan Africa that were extracted from from recent high-resolution satellite imagery (Ecopia.AI and Maxar Technologies, Inc. 2019–2021) (modal year = 2018, range = 2009 to 2019).

Census microdata. We used microdata samples from the 2010 Ghana census that were publicly available from IPUMS International (Minnesota Population Center 2019). This provides a count of the number of people from thousands of housing units with representative national coverage. Each household record is geo-tagged to geographic areas nested within larger regions. These level 1 and level 2 geographic boundaries (Fig. 1) were available for download from IPUMS International (Minnesota Population Center 2019).

There were no GPS coordinates associated with these household-level survey data for privacy reasons. The absence of specific location information is one of the primary challenges to overcome with the statistical model. Geospatial covariates that are usually fundamental for bottom-up statistical models (Wardrop et al. 2018,Leasure et al. 2020b) cannot be used here because we do not know the exact locations of the household surveys.

We used this to estimate the average number of people per household within sub-national geographic units. We also used these data to estimate the demographic composition of the populations in these areas by age and sex. We will assume that these parameters have not changed since 2010 when the census was conducted.

Population totals. We used population totals for level 3 administrative units that were projected to the year 2018 from the last census (WorldPop & CIESIN 2018a). We also used the level 3 administrative boundaries used for those projections (WorldPop & CIESIN 2018b) (Fig. 1).

We used this to estimate the average number of housing units per building. We used projections to the year 2018 to match the year of the building footprints.

Settlement type. We used the urban and rural classifications of enumeration areas from the 2010 Ghana census (Ghana Statistical Services 2010). We summarized the original three settlement types into urban and rural classes based on guidance from Ghana Statistical Services. We assumed that this was the same urban and rural classification reported in the census microdata from IPUMS International (Minnesota Population Center 2019).

Maps of settlement types, administrative boundaries, and IPUMS geographic regions (Fig. 1) were rasterized to a 100 m mastergrid matching the building footprints rasters (Dooley et al. 2020).

2.2 Statistical Model

We developed a Bayesian statistical model to estimate the average number of *people per housing unit* and the average number of *housing units per building* from these data.

2.2.1 Indexing

The indexing used throughout the model description (below) is as follows:

Households (h) contained within the IPUMS data. There were a total of 570,234 households from the Ghana census microdata and we used a representative 70% sample to fit our model (selected at random using the SAMPLE attribute of the IPUMS data). The remaining 30% was used for out-of-sample model validation.

Age-sex groups (k) that included 36 bins: under 1 year, 1 to 5 years, 80+ years, and 5 year intervals in between for both males and females.

Settlement type (t) for each enumeration area from the 2010 Ghana census. We summarized these as either urban (1) or rural (2).

Age-sex regions (r) were defined by the level 1 geographic boundaries from IPUMS (Minnesota Population Center 2019). We estimated the proportion of the population in each age-sex group for urban and rural areas within each of these 10 regions.

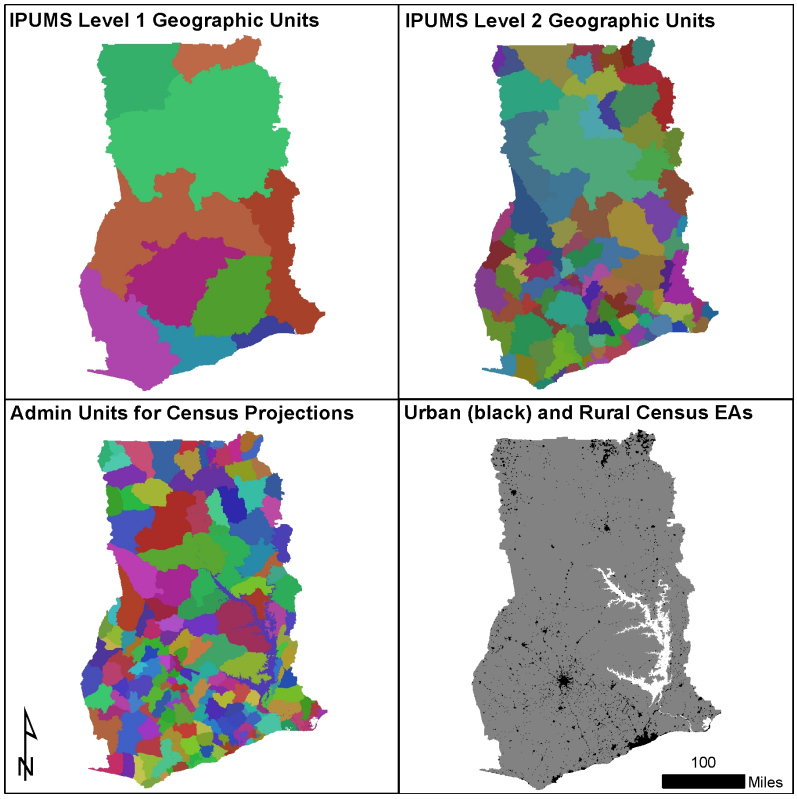


Figure 1: Geographic units used for modelling.

Household geographies (g) were defined by the level 2 geographic boundaries from IPUMS (Minnesota Population Center 2019). We estimated people per household for urban and rural areas within each of these 102 geographies.

Population units (u) were defined by the level 3 administrative boundaries from census projections (WorldPop & CIESIN 2018b). We estimated households per building for urban and rural areas within each of these 170 population units.

Locations (i) are 100 meter grid cells that were used for model predictions.

2.2.2 Population Estimates

To estimate the population for a given location i, we used the following model:

$$\begin{aligned} pop_i &\sim \text{Poisson}(bldg_i \times hpb_{t_i,u_i} \times pph_{t_i,g_i}) \\ pop_{i,k} &\sim \text{Multinomial}(pop_i, \theta_{t_i,r_i,k}) \end{aligned} \tag{1}$$

pop_i is the total population estimated for location i, and $pop_{i,k}$ is the number of people within age-sex group k. $bldg_i$ is the observed number of building footprints at this location. hpb and pph are estimated parameters measuring households per building and people per household, respectively. In this portion of the model, the only observed data are the building counts $bldg_i$, settlement type t_i, and the three spatial unit IDs (g_i, u_i, and r_i). Note that we are assuming that the count of building footprints is equivalent to the true count of buildings.

Estimating the parameters $hpb_{t,u}$, $pph_{t,g}$, and $\theta_{t,r,k}$ is the focus of the statistical model developed below.

2.2.3 People per Household (pph)

Single-person households are more common in the Ghana microdata than could be accounted for by a simple Poison distribution to model household sizes, so we used a Hurdle model (Stan Development Team 2019) to account for

the one-inflated distribution. This involves two processes:

1. $\alpha_{t,g}$ the probability that a household has only a single-person , and
2. $\lambda_{t,g}$ the number of additional people in multi-person households other than the head of household.

These two parameters are the basis for estimating people per household $pph_{t,g}$ by type and household geographic unit. The Ghana census microdata (IPUMS 2020) provided the data that we needed to fit this part of the model. This included a count of people and their ages from a representative sample of households across the country.

We used a beta-binomial model to estimate the probability of a single-person household $\alpha_{t,g}$ in urban and rural settlement types t in each geographic area g:

$$\begin{aligned} single_{t,g} &\sim \mathsf{Binomial}(hh_{t,g}, \alpha_{t,g}) \\ \alpha_{t,g} &\sim \mathsf{Beta}(\frac{\bar{\alpha}_t}{\tilde{\alpha}}, \frac{1-\bar{\alpha}_t}{\tilde{\alpha}}) \end{aligned} \tag{2}$$

where $hh_{t,g}$ is the total number of households surveyed from settlement type t in geography g, and $single_{t,g}$ is the total number of single-person households surveyed. $\alpha_{t,g}$ is the probability that a household contains only a single person. $\bar{\alpha}_t$ is the mean of $\alpha_{t,g}$ among geographies within settlement type t and $\tilde{\alpha}$ quantifies variance among geographies.

We used a Poisson model with LogNormal overdispersion to estimate the number of people in multi-person households $\lambda_{t,g}$ (excluding the head of household):

$$\begin{aligned} N_h - 1 &\sim \mathsf{Poisson}(\lambda_h)T(1,) \\ \lambda_h &\sim \mathsf{LogNormal}(\mu_{t,g}, \sigma_{t,g}) \end{aligned} \tag{3}$$

where N_h is the observed number of people in each multi-person household h from the IPUMS census microdata and λ_h is the expected count of household members other than the head of household. We used $N_h - 1$ as the response variable so we could use zero-inflation to account for one-inflated distributions of N_h. The Poisson distribution is truncated below 1 (notated as T) because the beta-binomial model (Eq. (2)) already accounts for single-person households. $\mu_{t,g}$ is the mean of λ_h (on the log scale) among households from settlement type t in geography g, and $\sigma_{t,g}$ is the variance (on the log scale) among households in settlement type t. We modelled $\mu_{t,g}$ and $\sigma_{t,g}$ hierarchically to share information among geographies:

$$\begin{aligned} \mu_{t,g} &\sim \mathsf{Normal}(\bar{\mu}_t, \tilde{\mu}) \\ \sigma_{t,g} &\sim \mathsf{Half\text{-}Normal}(\bar{\sigma}_t, \tilde{\sigma}) \end{aligned} \tag{4}$$

where $\bar{\mu}_t$ is the mean value of $\mu_{t,g}$ among geographies within settlement type t, and $\tilde{\mu}$ is the variance among geographies in both settlement types combined. We used these parameters to estimate a version of λ that was not constrained by household-level observations:

$$\hat{\lambda}_{t,g} \sim \mathsf{LogNormal}(\mu_{t,g}, \sigma_{t,g}) \tag{5}$$

When we say that $\hat{\lambda}_{t,g}$ was not constrained by household-level observations, we mean that it is constant among households within a settlement type in a given geography. This is neccessary to derive the expected values (i.e. the means) of people per household to use for population predictions:

$$E(pph_{t,g}) = 1 + (1 - \alpha_{t,g})\hat{\lambda}_{t,g} \tag{6}$$

The priors that we used for parameters in this part of the model were:

$$
\begin{aligned}
\bar{\alpha_t} &\sim \mathsf{Uniform}(0, 1) \\
\tilde{\alpha} &\sim \mathsf{Uniform}(0, 1) \\
\bar{\mu_t} &\sim \mathsf{Normal}(0, 5) \\
\tilde{\mu} &\sim \mathsf{Uniform}(0, 5) \\
\bar{\sigma_t} &\sim \mathsf{Uniform}(0, 5) \\
\tilde{\sigma} &\sim \mathsf{Uniform}(0, 5)
\end{aligned}
\tag{7}
$$

Our aim was to use relatively uninformative priors that provided information about the range of possible values but were not influential on parameter estimates relative to the data.

Just for comparison purposes, another way to write the Hurdle model from Eqs. (2) and (3) is:

$$
p(N_h - 1 \mid \alpha_{t,g}, \lambda_h) = \begin{cases} \alpha_{t,g} & \text{if } N_h - 1 = 0, \text{ and} \\ (1 - \alpha_{t,g}) \dfrac{\mathsf{Poisson}(N_h - 1 \mid \lambda_h)}{1 - \mathsf{PoissonCDF}(0 \mid \lambda_h)} & \text{if } N_h - 1 > 0, \end{cases}
\tag{8}
$$

See the Stan User's Guide (Stan Development Team 2019) for more details about the Hurdle model specification.

2.2.4 Demographic Groups

We used a multinomial-Dirichlet model to account for demographic structure in the population:

$$
\begin{aligned}
M_{t,r,k} &\sim \mathsf{Multinomial}(\dot{M}_{t,r}, \theta_{t,r,k}) \\
\theta_{t,r,k} &\sim \mathsf{Dirichlet}(rep(1, K))
\end{aligned}
\tag{9}
$$

where $\dot{M}_{t,r}$ is the total number of survey respondents from our IPUMS sample in settlement type t within region r, and $M_{t,r,k}$ is the number of respondents within age-sex group k observed in the census microdata. In this model, the age-sex structure $\theta_{t,r,k}$ (i.e. proportion of population in each group) is a parameter that is estimated independently for each region and settlement type. The Dirichlet prior is an uninformative flat prior. K is the total number of age-sex groups (i.e. $K = 36$). The Dirichlet distribution produces a vector of probabilities (i.e. one for each age-sex group) that must sum to one (i.e. to represent the total population).

2.2.5 Households per Building (hpb)

For each population unit u we know the total count of buildings and the estimated total population from census projections. From the previous model components we also have an estimate of the average number of people per household for household geographies g and settlement types t within the population unit. Our goal now is to use these pieces of information to estimate the average number of housing units per building for urban and rural areas within each population unit u. We accomplished this using the following model:

$$
\begin{aligned}
\dot{N}_u &\sim \mathsf{LogNormal}(log(\bar{N}_u), \tilde{N}_u) \\
\bar{N}_u &= \sum_{t=1}^{T} \sum_{g=1}^{G} (B_{t,u,g} \times hpb_{t,u} \times pph_{t,g})
\end{aligned}
\tag{10}
$$

where $\dot{N}_u$ are the estimated total population sizes for unit u that were provided as input data from census projections (WorldPop & CIESIN 2018a). $B_{t,u,g}$ is the observed number of building footprints within the intersection of geography g, population unit u, and settlement type t. We modelled $hpb_{t,u}$ hierarchically using a log-normal distribution to share information among population units:

$$hpb_{t,u} \sim \mathsf{LogNormal}(\bar{hpb}_t, \tilde{hpb}) \tag{11}$$

where $\bar{hpb}_t$ is the mean of $hpb_{t,u}$ among population units within settlement type t, and $\tilde{hpb}$ is the variance among population units including both settlement types. These variance terms did not differ significantly among settlement types, so estimating a single variance parameter produced similar results to estimating them seperately.

2.2.6 Measurement Error in Census Projections

We treated the parameter $\tilde{N}_u$ from Eq. (10) as measurement error in the projected census totals $\dot{N}_u$ that were used as input data. We provided estimates of measurement error as an informative prior:

$$\tilde{N}_u \sim \mathsf{LogNormal}(log(0.15), 0.5) \tag{12}$$

Because we did not have information about the uncertainty of the census projections, we designed this prior to be as vague as possible while still resulting in model convergence. The mean of log(0.15) represents our prior belief that the projected census totals were within about 25% of the true population totals, on average across population units u. The standard deviation of 0.5 represents our prior belief that there may have been significantly more measurement error in some population units (e.g. 100% or more). Because the model is estimating this parameter for every population unit, the model has the freedom to identify which units have more or less measurement error, based on information from the rest of the model.

2.3 Model Implementation and Diagnostics

We implemented this Bayesian model using the R statistical programming language (R Core Team 2020) and the rstan package (Stan Development Team 2020). The model code and input data are provided as supplementary files that are described in Appendix A. Convergence of MCMC chains (i.e. Markov chain Monte Carlo) was evaluated using the Rhat metric from the Stan software (Stan Development Team 2019, Stan Development Team (2020)). We used four MCMC chains with a warmup period of 250 iterations followed by a sampling period of 2500 MCMC iterations per chain.

We have three types of population data that we can use to evaluate model fit: people per household and people per age-sex group from the household-level census microdata (Minnesota Population Center 2019) and estimates of total population per region from projected census totals (WorldPop & CIESIN 2018b).

In general, we assessed model fit using model residuals `(observed - predicted)` to calculate bias `mean(residuals)`, imprecision `sd(residuals)`, inaccuracy `mean(abs(residuals))`, and r-squared percent variance explained (squared Pearson correlation coefficient). The mean values from posterior predictions were used for these assessments. We assessed the model's prediction intervals by calculating the proportion of out-of-sample observations that were within the prediction intervals. If the prediction intervals were robust, we expected about 95% of out-of-sample observations to fall inside the 95% prediction intervals and 50% of observations fall inside of 50% prediction intervals.

We assessed model fit at the household-level using the 30% out-of-sample subset from the census microdata (Minnesota Population Center 2019). We analyzed residuals for the following parameters:

- $\alpha_{t,g}$ probability of a single-person household,
- $\lambda_{t,g}$ additional household members in multi-person households,
- $pph_{t,g}$ number of people per household,
- $\theta_{t,r,k}$ proportion of the population in each age-sex group, and
- $\dot{N}_u$ total population per population unit u.

We assessed model fit at the region-level using in-sample data from the projected census totals (WorldPop & CIESIN 2018a). We generated a posterior predictions for the total population in unit u as:

$$\hat{\dot{N}}_u \sim \mathsf{LogNormal}(\bar{\dot{N}}_u, \tilde{\dot{N}}_u) \tag{13}$$

Spatial units u where there was a large difference between the modelled population total $\hat{\dot{N}}$ and the projected census totals $\dot{N}$ likely indicates that census projections were inaccurate in those units, but it could also indicate a lack of model fit for that location (i.e. biased estimates of $hpb_{t,u}$ or $pph_{t,g}$).

3 Results

The model reached full convergence for all parameters based on the Rhat metric used by Stan and there were no additional warnings about instabilities of means, medians, or tails of posteriors.

Full model results can be downloaded from the WorldPop Open Population Repository (Leasure & Tatem 2020). This included 100 m gridded estimates with national coverage for:

1. The number of people,
2. The number of people in each age-sex group,
3. The number of households,
4. The average number of people per household,
5. The average number of households per building footprint, and
6. The average number of people per building footprint.

Posterior estimates for total number of people and people per age-sex group can be explored on an interactive map using the wopr R package and the woprVision web application (Fig. 2) (Leasure, Bondarenko & Tatem 2020). It is not possible to validate the model predictions at the 100 m grid cell level because we do not have geolocated population data. The posterior predictions have wide confidence intervals that accurately reflect this uncertainty.

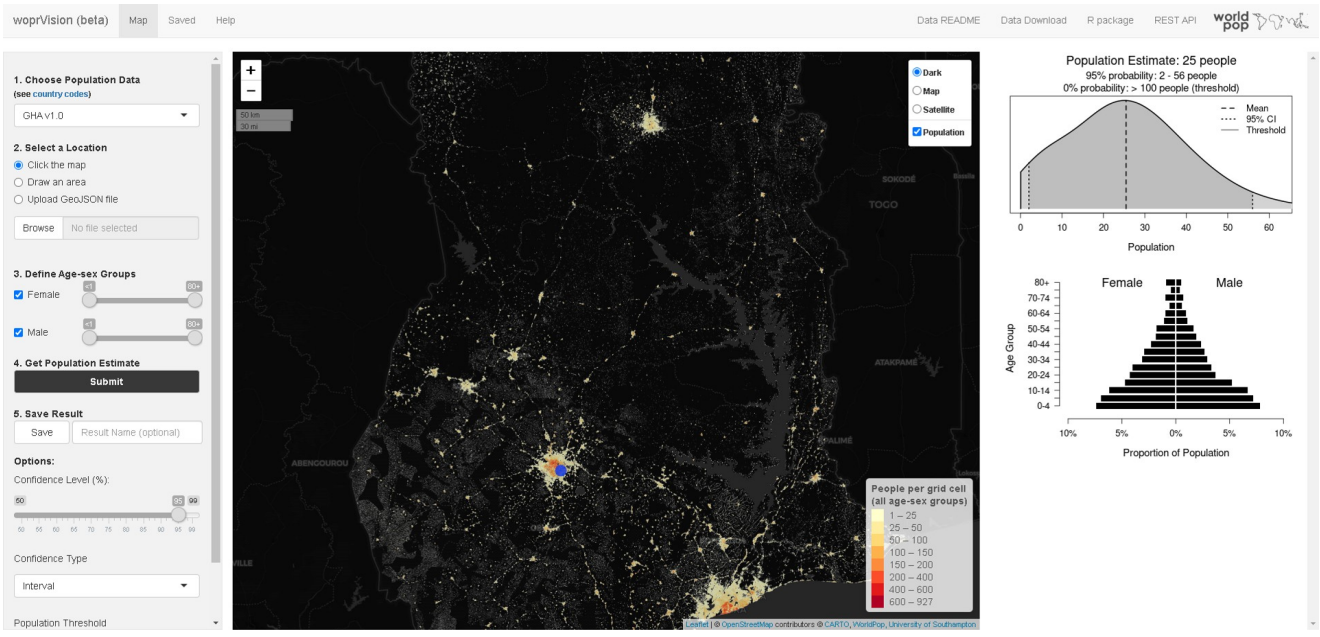


Figure 2: woprVision web application providing access to posterior predictions on an interactive web map (https://apps.worldpop.org/woprVision; select data 'GHA v1.0').

3.1 People per household

Analysis of residuals indicated that estimates of people per household were relatively unbiased (-6.4%) but predictoins were very imprecise (76%) for individual households (r-squared = 0.05). This was expected due to the

lack of location information from the household data. Despite imprecise household-specific predictions, the overall distributions of predicted household sizes were representative of household sizes within geographic units and settlement types (Fig. 3). The prediction intervals for people per household were robust with 96.8% of out-of-sample data falling within the 95% prediction interval and 69% falling within the 50% prediction interval.

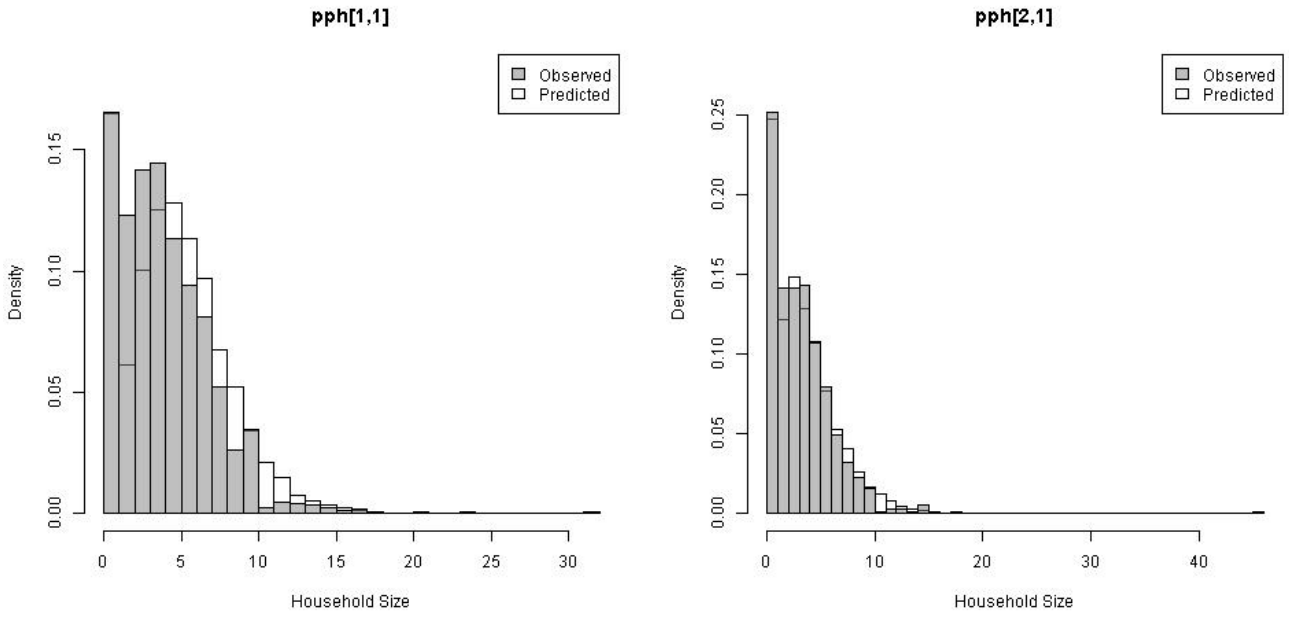


Figure 3: Distribution of people per household in geographic unit no. 1 for urban areas (left; t=1, g=1) and rural areas (right; t=2, g=1).

Estimates of $\alpha_{t,g}$ (bias = 1.3%, imprecision = 19.2%) and $\lambda_{t,g}$ (bias = -1.3%, imprecision = 8.3%) were much more accurate because they were estimated for larger spatial scales (geographic units g) with r-squared values of 0.88 and 0.93, respectively (Fig. 4). Remember that $\alpha_{t,g}$ is the probability that a household contains only a single person and $\lambda_{t,g}$ is the average number of people in a household other than the head of household. We did not assess the coverage of the prediction intervals for each of these parameters individually because they are component parts of a mixture distribution representing people per household, which was assessed above.

3.2 Age-sex structure

Estimates of $\theta_{t,r,k}$ (i.e. proportion of population in each age-sex group) had an r-squared of 0.98 compared to out-of-sample data (Fig. 5; bias = -6.6%, imprecision = 18%). The slight negative bias was particularly apparent for estimates of adult males in some geographic units (see Appendix B). The 95% prediction intervals contained only 48% of out-of-sample data and the 50% prediction intervals contained only 21% of out-of-sample data. The underestimated prediction intervals may have resulted from the strong constraints of the Dirichlet distribution requiring the proportions to sum to one.

The population pyramids that resulted from the estimates of theta differed slightly between urban areas and rural areas (Fig. 6). Urban areas had a wider base indicating a higher proportion of children. Within settlement types, there were slight differences in population pyramids among geographic units (see Appendix B).

Figures comparing model estimates of people per household and age-sex structure to out-of-sample observations (like Figs. 3 and 6) are provided in Appendix A for all geographic units.

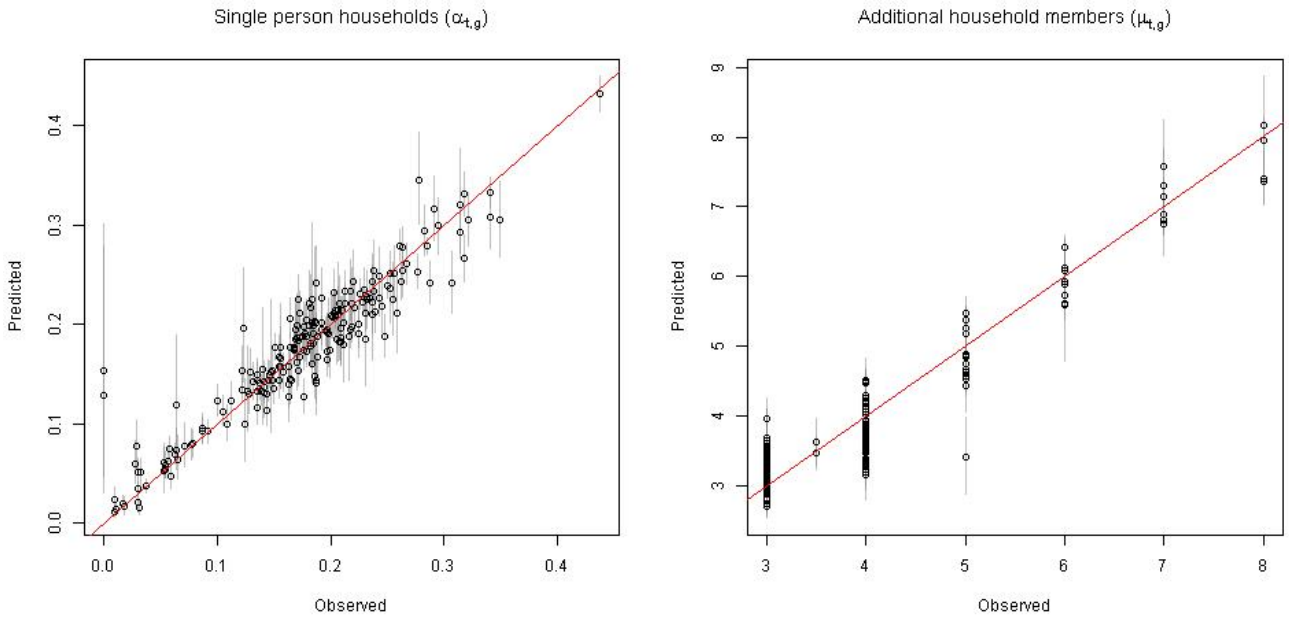


Figure 4: Observed versus predicted parameters for estimating people per household.

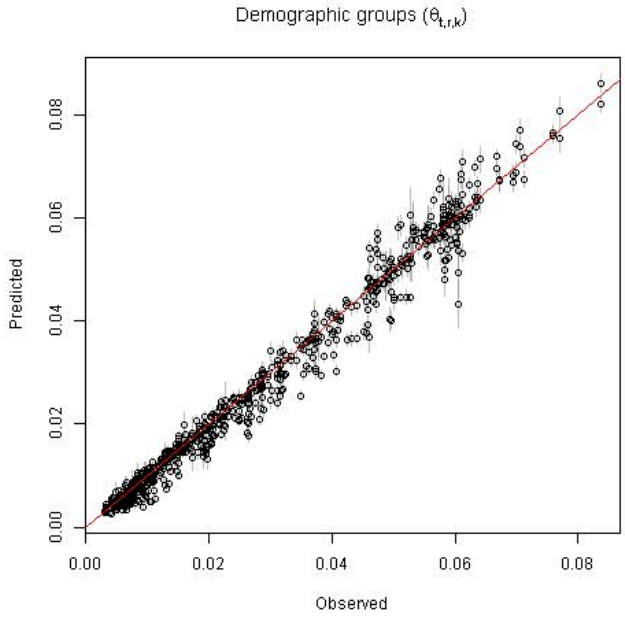


Figure 5: Observed versus predicted proportion of the population in each age-sex group. The plot shows all age-sex groups across all regions.

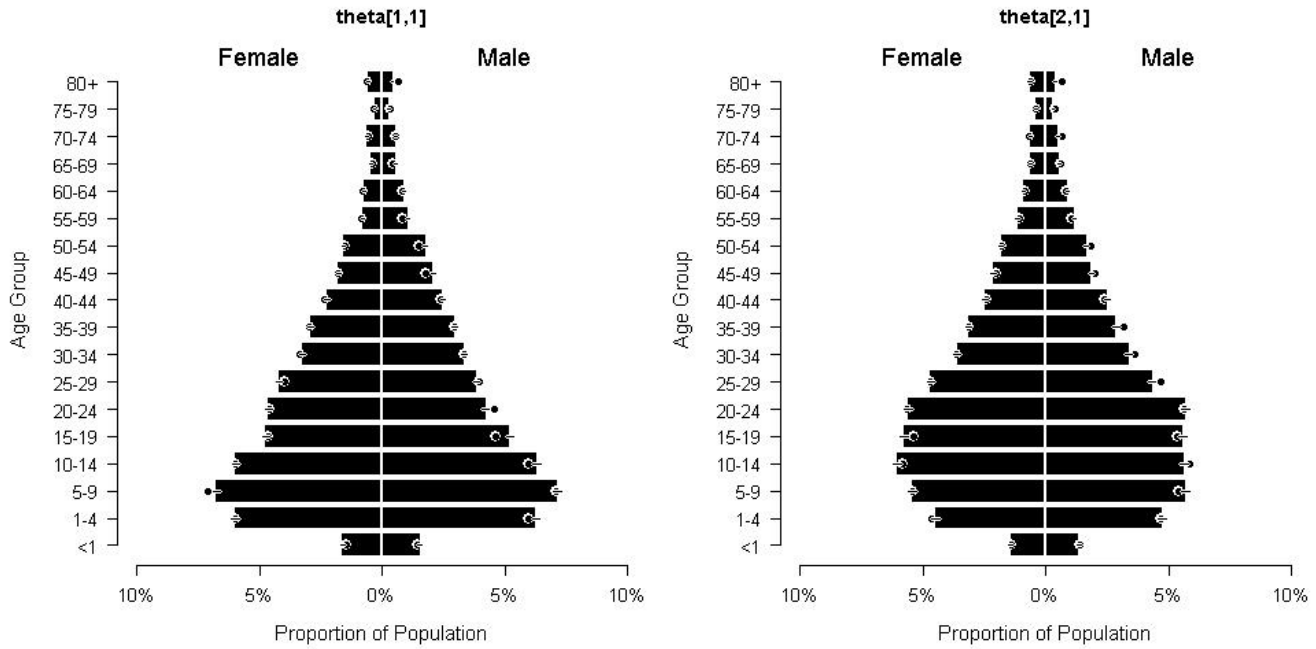


Figure 6: Population pyramid in region no. 1 for urban areas (left, t=1, r=1) and rural areas (right, t=2, r=1). Dots are observed out-of-sample data, while bars and their credible intervals are posterior predictions from the model.

3.3 Census projections

Model estimates of total population in each population unit u were comparable to the projected census totals used as input data for most areas (r-squared = 0.98), but there were significant divergences in two units (Fig. 7) where the census projections were likely significant underestimates that did not account for urban growth into these suburban/rural areas. One of the areas was on the north side of the city of Kumasi and the other was west of the city of Tamale (see results at https://apps.worldpop.org/woprVision; select data "GHA v1.0").

4 Discussion

The model is a hybrid between commonly used top-down and bottom-up approaches (Wardrop et al. 2018) that are often used to map populations when complete census results are not available. It is like bottom-up approaches because it uses household-level survey data that do not have full coverage of the country. It is like top-down approaches because it uses projected census totals to constrain population estimates. Unlike other top-down approaches (Stevens et al. 2015), this model enforces a "soft constraint" in which the population totals can deviate significantly from the census projections provided as input data if the weight of evidence from the rest of the model suggests the projections are inaccurate. This approach is unlike top-down and bottom-up approaches because it cannot use high-resolution geospatial covariates other than the count of buildings in each 100 m pixel.

This modelling approach produces high resolution gridded population estimates from publicly available census microdata that have been standardized across countries by IPUMS International (Minnesota Population Center 2019). The building footprints that are the other critical piece of data have been produced for 51 countries in sub-Saharan Africa (Dooley et al. 2020,Ecopia.AI and Maxar Technologies, Inc. 2019–2021). This provides an opportunity to produce rapid population estimates for countries where both data sets are available. We achieved this by estimating the average number of *households per building* and the average number of *people per household* for urban and rural sub-national geographic units using a hierarchical Bayesian modelling approach. Importantly, the method accounts for uncertainty which can be significant for approaches like this that do not have detailed

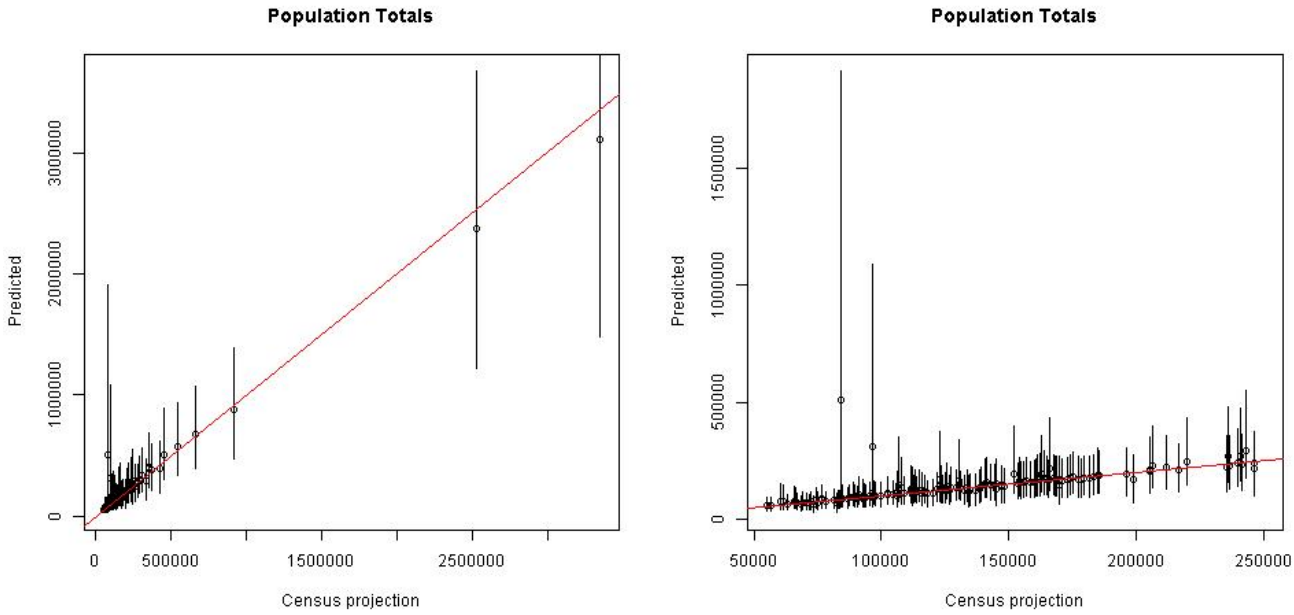


Figure 7: Census projections versus model predictions of total population sizes in 170 spatial units. The panel on the right is zoomed in to units with less than 250,000 people.

spatial information associated with the population data, which is the usual situation when using publicly available data.

This method cannot map variation in people per household or households per building with precision at the 100 m spatial scale because specific locations of households are not available for the census microdata. The model estimates these parameters for urban and rural settlements in over 100 geographic units across the country, but it does not attempt to map variation within those units. It accounts for that variation as statistical uncertainty which results in the population estimates having fairly wide credible intervals. Because we used data from the 2010 Ghana census to estimate the number of people per household and the proportion of the population in each age-sex group for urban and rural settlements in various regions, we have assumed that the patterns of people per household and age-sex structure have not changed in these areas since 2010. This was a necessary assumption in the absence of more recent survey data, but the model can be updated when newer data become available.

We suggest labelling the population estimates as 2018 because this was the most common satellite image acquisition year (Fig. 8) for imagery used by Maxar Technologies and Ecopia.AI (Ecopia.AI and Maxar Technologies, Inc. 2019–2021) to derive the building footprints. The year of satellite images varied across the country from 2009 to 2019, depending on the exact location. Refer to the raster GHA_buildings_v1_1_imagery_year.tif (Dooley et al. 2020) for a map of the imagery years of building footprints. We assumed that the building footprints represented all potentially residential buildings. We did not account for buildings that were missing from the building footprints or that have since been removed. We did not have quantitative measures of building footprint accuracy available to us.

We identified a few areas for improvement in this model. The credible intervals for the age-sex structure are not capturing the full range of variation. One potential solution may be to apply a hierarchical Dirichlet prior to constrain the population pyramids to share information among regions. The estimated distributions of people per household appeared not to fit the out-of-sample data for a few geographic units and this warrants further investigation, particularly to understand the effects of outliers on the one-inflated Hurdle model. It would also be helpful to conduct a formal sensitivity analysis to assess the influence of the informative prior that we used to quantify measurement error in the census projections.

In addition to these next steps to improve the method and better understand its nuances, it will be important to apply it across countries to understand how generalizable it is. Our intention for this approach is to apply it

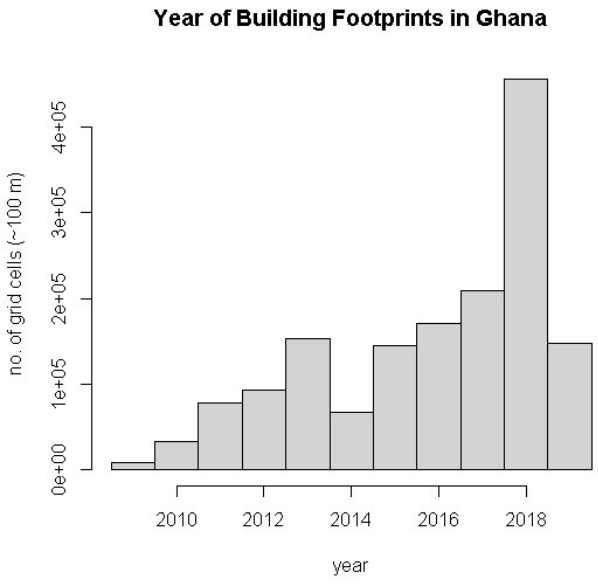


Figure 8: Distribution across Ghana of the year of building footprints (Ecopia.AI and Maxar Technologies 2020, Dooley et al. 2020).

rapidly as a gap-filling measure while efforts are underway to acquire geolocated population data for population modelling to support development projects, planning for programs, and census preperations. The model's estimates of people per household and households per building may also serve as a guide for setting these parameters in the peanutButter application (Leasure et al. 2020a) for countries where census microdata or other household surveys are not publically available.

5 Contributing

This report and the methods it describes were produced by Doug Leasure from the WorldPop Research Group at the University of Southampton with oversight from Andy Tatem. The work was funded by the Bill and Melinda Gates Foundation to provide critical spatial data and population estimates to support polio surveillance and eradication (INV-002697). We want to acknowledge Vince Seaman from the Bill and Melinda Gates Foundation and Kebba Touray from the World Health Organization (AFRO-Polio GIS centre) for facilitating access to data for this project. We want to thank Edith Darin, Chris Jochem, and Attila Lazar for thoughtful internal reviews that helped improve the work.

6 Suggested Citation

Leasure DR, Tatem AJ. 2020. A Bayesian approach to produce 100 m gridded population estimates using census microdata and recent building footprints. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00686

7 License

8 References

Dooley CA, Boo G, Leasure DR, Tatem AJ. 2020. Gridded maps of building patterns throughout sub-saharan africa, version 1.1. doi:10.5258/SOTON/WP00677.

Ecopia.AI and Maxar Technologies, Inc. 2019–2021. *Digitize Africa Data.* Ecopia.AI and Maxar Technologies, Inc.

Ghana Statistical Services. 2010. *Census enumeration area boundaries and urban/rural classification.* Accra, Ghana: Ghana Statistical Services.

Leasure DR, Tatem AJ. 2020. *Bayesian gridded population estimates for ghana 2018, version 1.0.* WorldPop, University of Southampton. doi:10.5258/SOTON/WP00680. https://wopr.worldpop.org/?GHA/Population/v1.0.

Leasure DR, Bondarenko M, Tatem AJ. 2020. *wopr: An R package to query the WorldPop Open Population Repository, version 0.3.4.* WorldPop, University of Southampton. doi:10.5258/SOTON/WP00679. https://apps.worldpop.org/woprVision.

Leasure DR, Dooley CA, Bondarenko M, Tatem AJ. 2020a. peanutButter: An R package to produce rapid-response gridded population estimates from building footprints, version 0.3.0. doi:10.5258/SOTON/WP00681.

Leasure DR, Jochem WC, Weber EM, Seaman V, Tatem AJ. 2020b. National population mapping from sparse survey data: A hierarchical bayesian modeling framework to account for uncertainty. *Proceedings of the National Academy of Sciences.* doi:10.1073/pnas.1913050117.

Minnesota Population Center. 2019. *Integrated Public Use Microdata Series, International: Version 7.2 [Ghana 2010].* Minneapolis, MN: IPUMS.

R Core Team. 2020. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Stan Development Team. 2019. *Stan User's Guide.* https://mc-stan.org/docs/2_23/stan-users-guide.

Stan Development Team. 2020. *RStan: the R interface to Stan. R package version 2.19.3.* http://mc-stan.org/.

Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one* 10:e0107042. doi:10.1371/journal.pone.0107042.

Wardrop NA, Jochem WC, Bird TJ, Chamberlain HR, Clarke D, Kerr D, Bengtsson L, Juran S, Seaman V, Tatem AJ. 2018. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences* 115:3529–3537. doi:10.1073/pnas.1715305115.

WorldPop, CIESIN. 2018a. *Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076).* WorldPop, University of Southampton. doi:10.5258/SOTON/WP00651.

WorldPop, CIESIN. 2018b. *Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076).* WorldPop, University of Southampton. ftp://ftp.worldpop.org/GIS/Population/Global_2000_2020/CensusTables/.

9 Appendix A: Supplementary Data

9.1 Model Data

The attached file `data.rds` is an R object that can be read into the R environment with the code `dat <- readRDS('data.rds')`. It is a list object with an element for each observed variable in the statistical model.

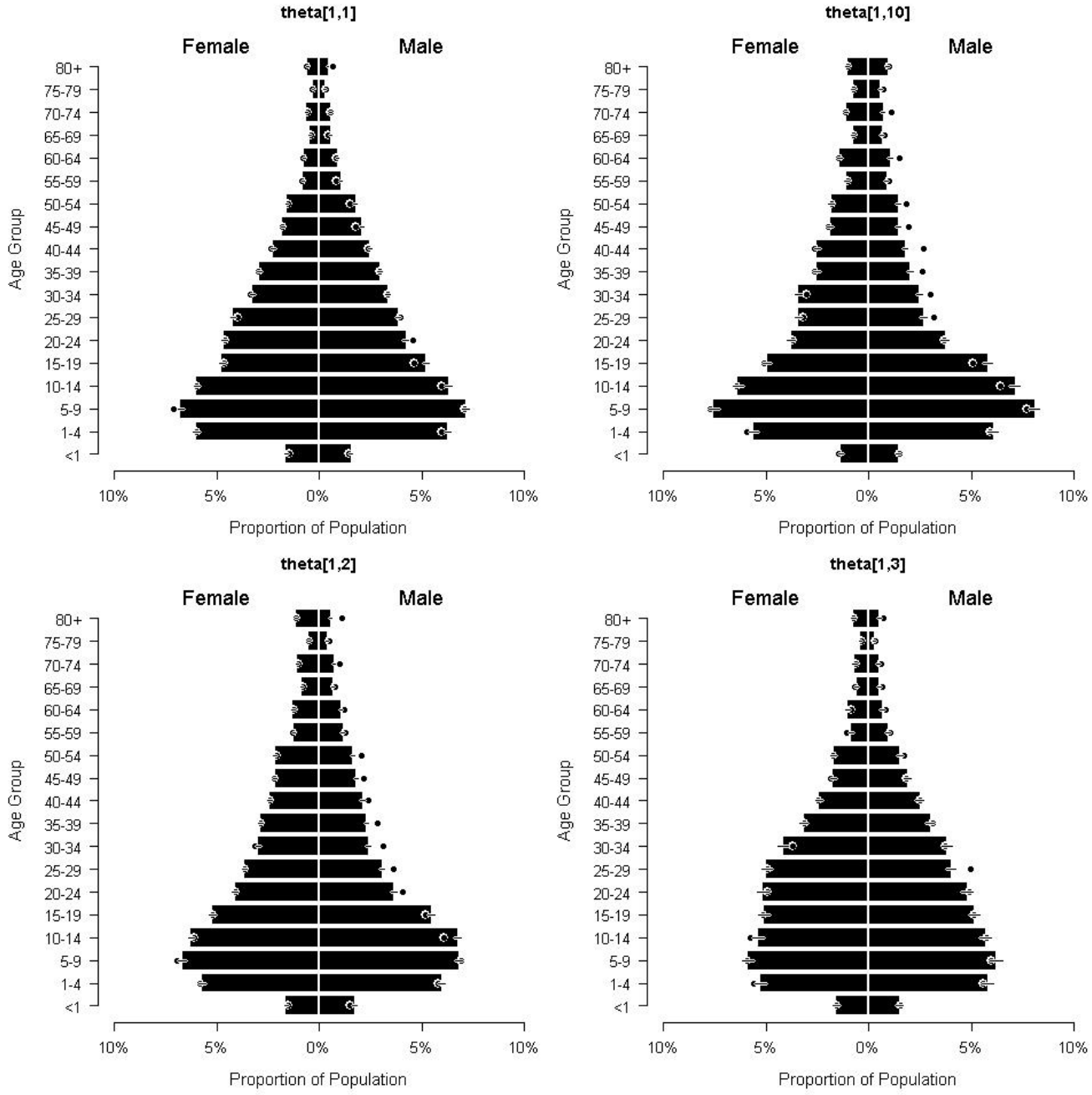
We have replaced the counts of people per household (N) and total population in each age-sex group (M) with NAs due license restrictions from IPUMS International that prohibit redistribution of these data to protect the privacy of individuals represented in the data set. You must request access to the data directly from IPUMS International (https://international.ipums.org/international/).

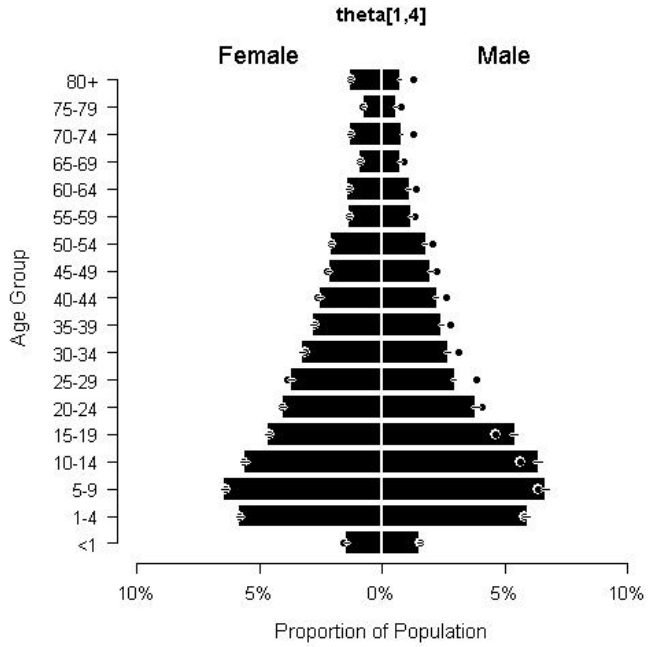
9.2 Model Code

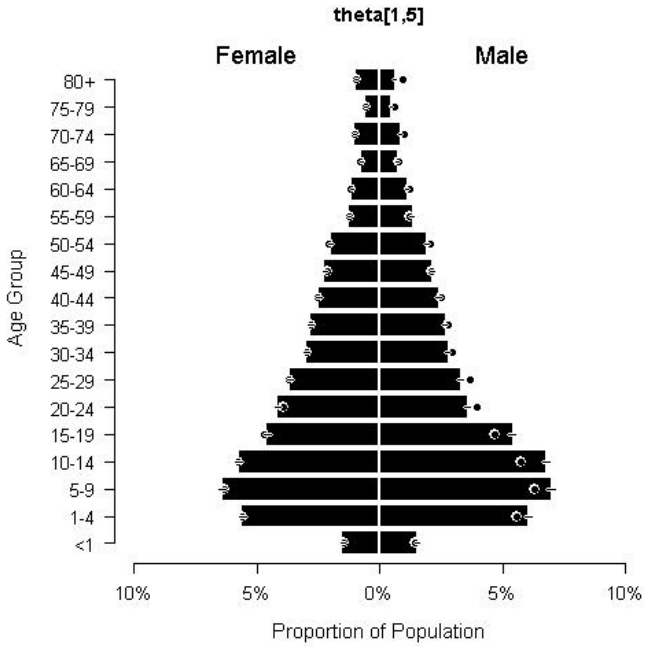
The attached file `model.stan` is a text file that contains the statistical model in the Stan programming language (Stan Development Team 2019).

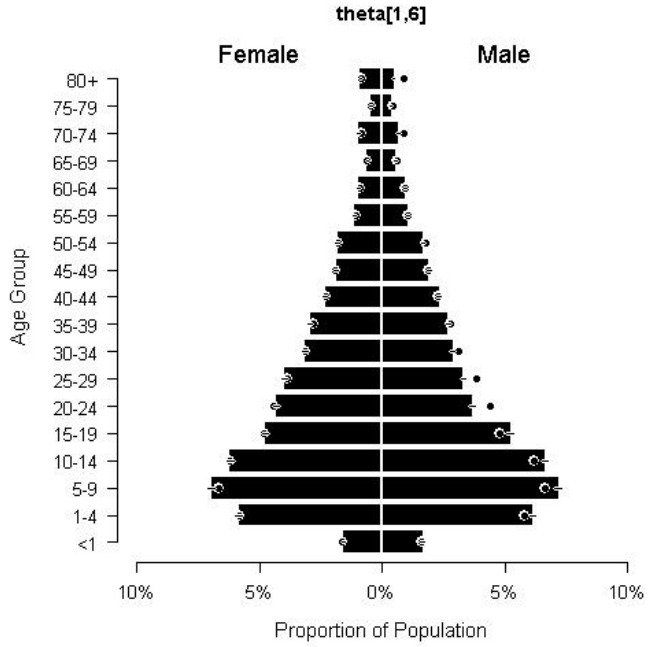
10 Appendix B: Supplementary Plots

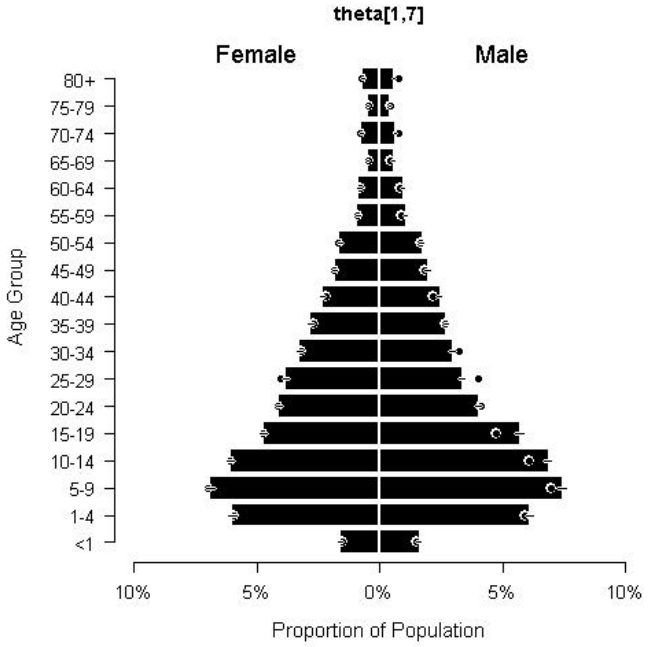
10.1 Age-sex structure

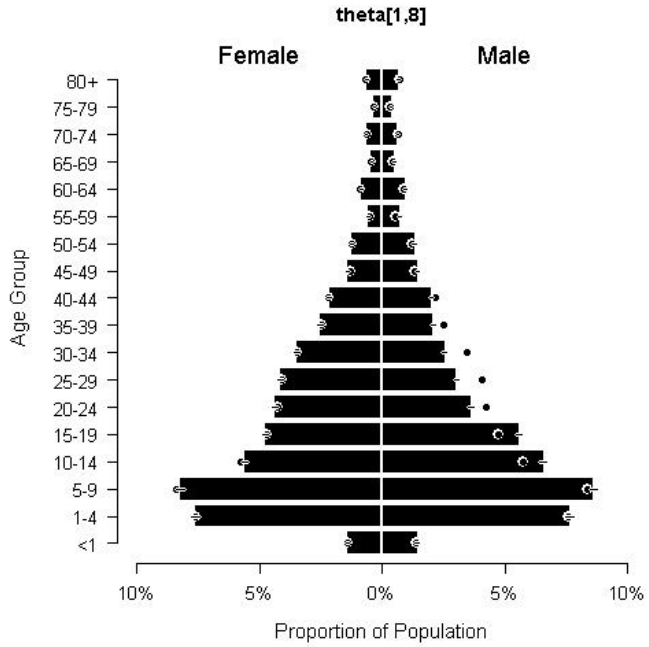


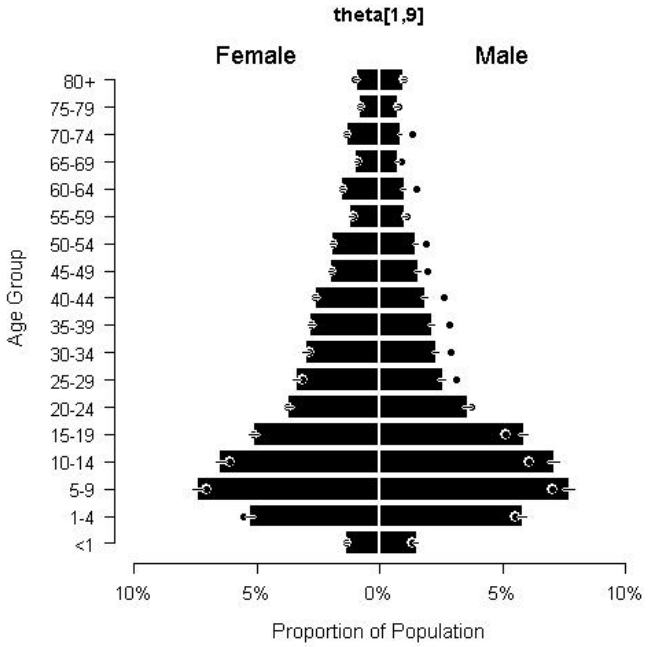

theta[1,4]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

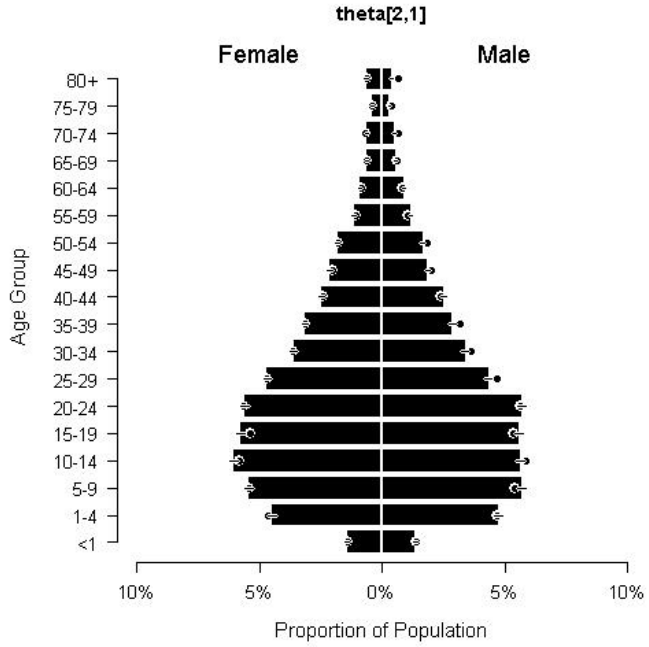

theta[1,5]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

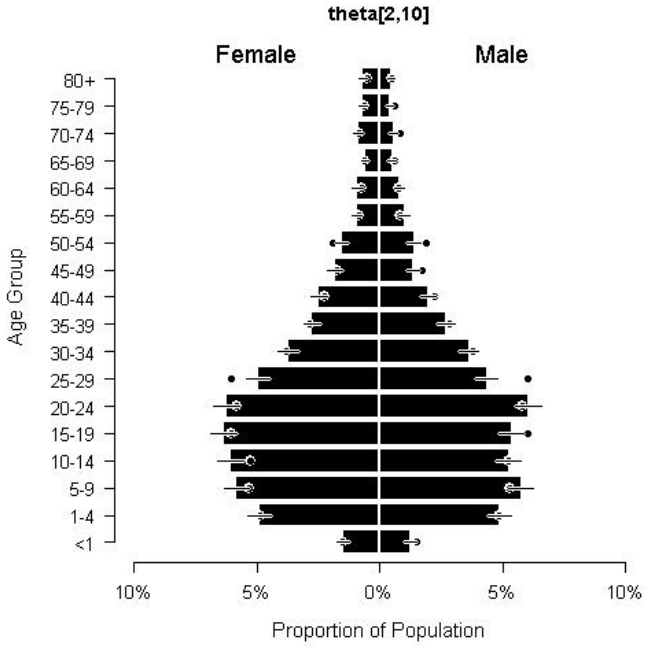

theta[1,6]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

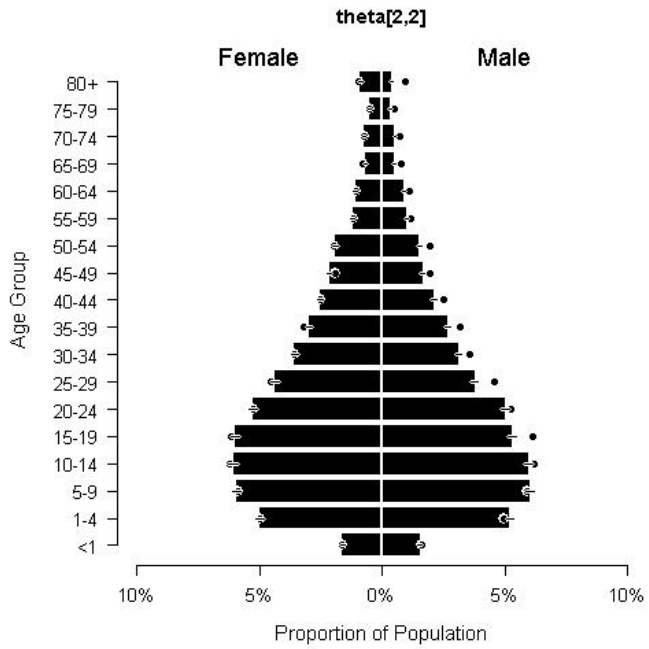

theta[1,7]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

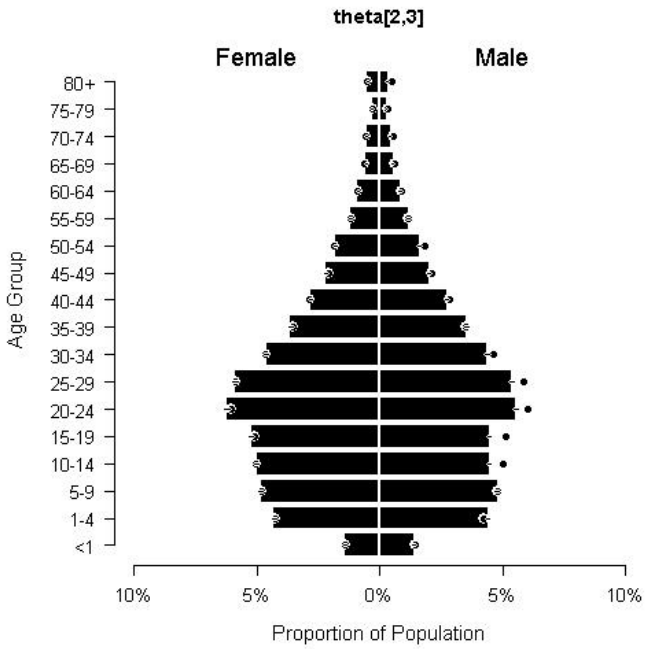

theta[1,8]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

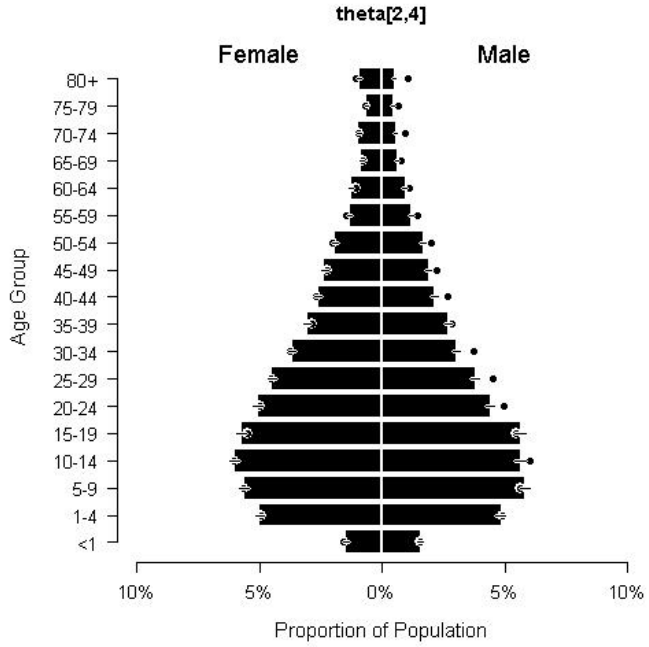

theta[1,9]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

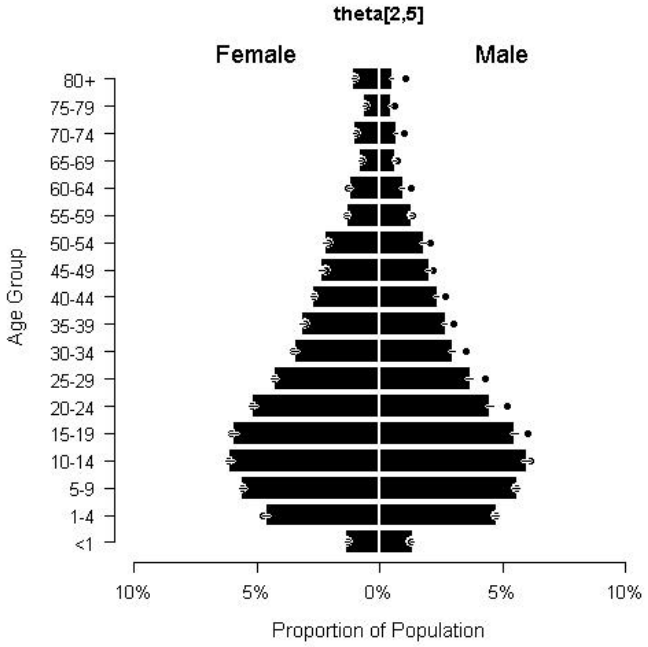

theta[2,1]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

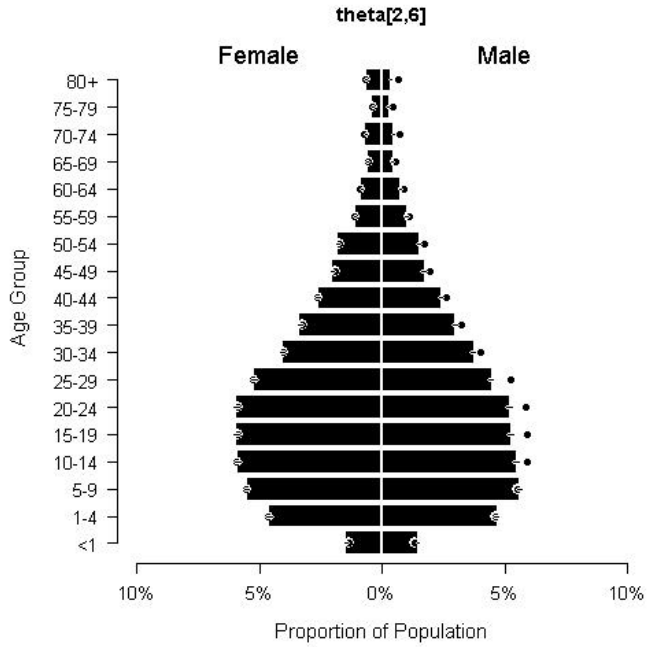

theta[2,10]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

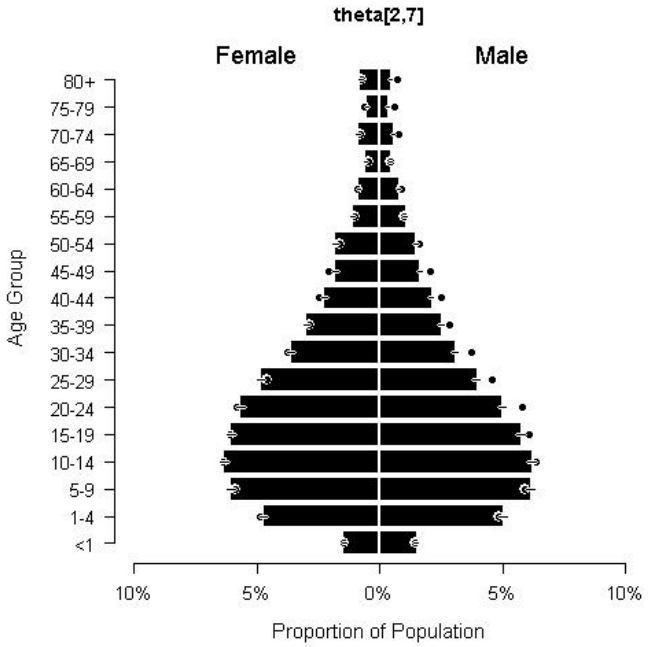

theta[2,2]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

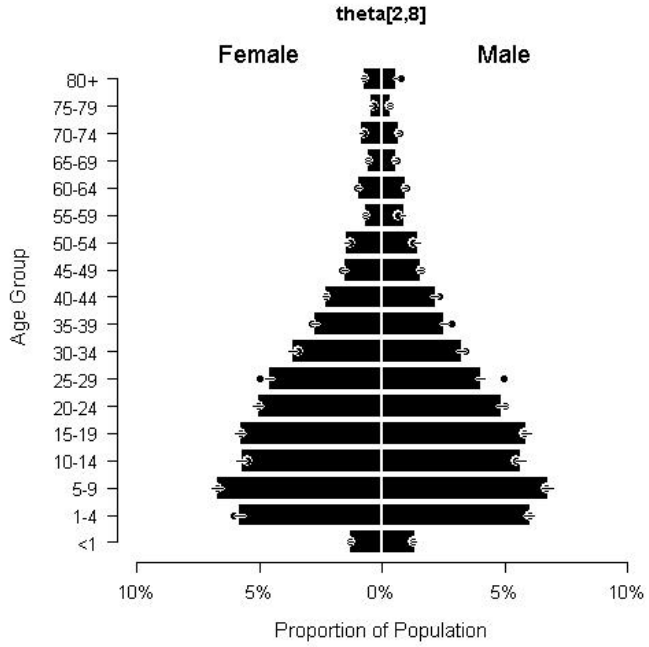

theta[2,3]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

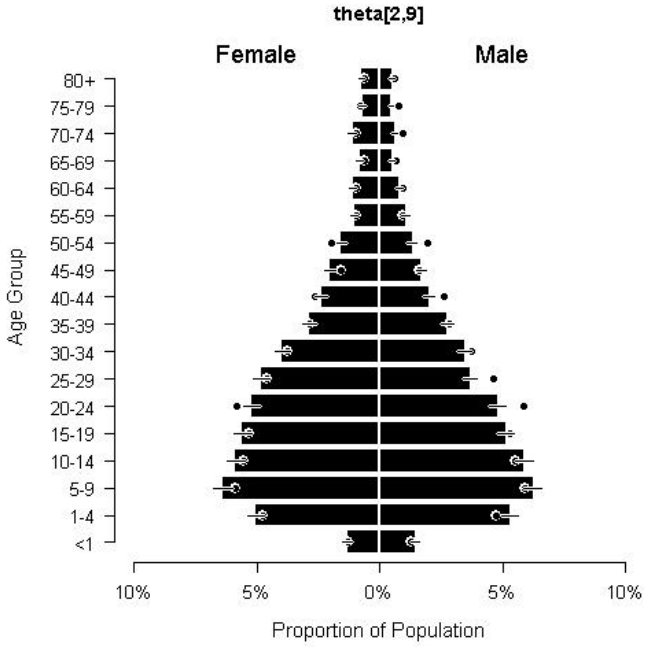

theta[2,4]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population


theta[2,5]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

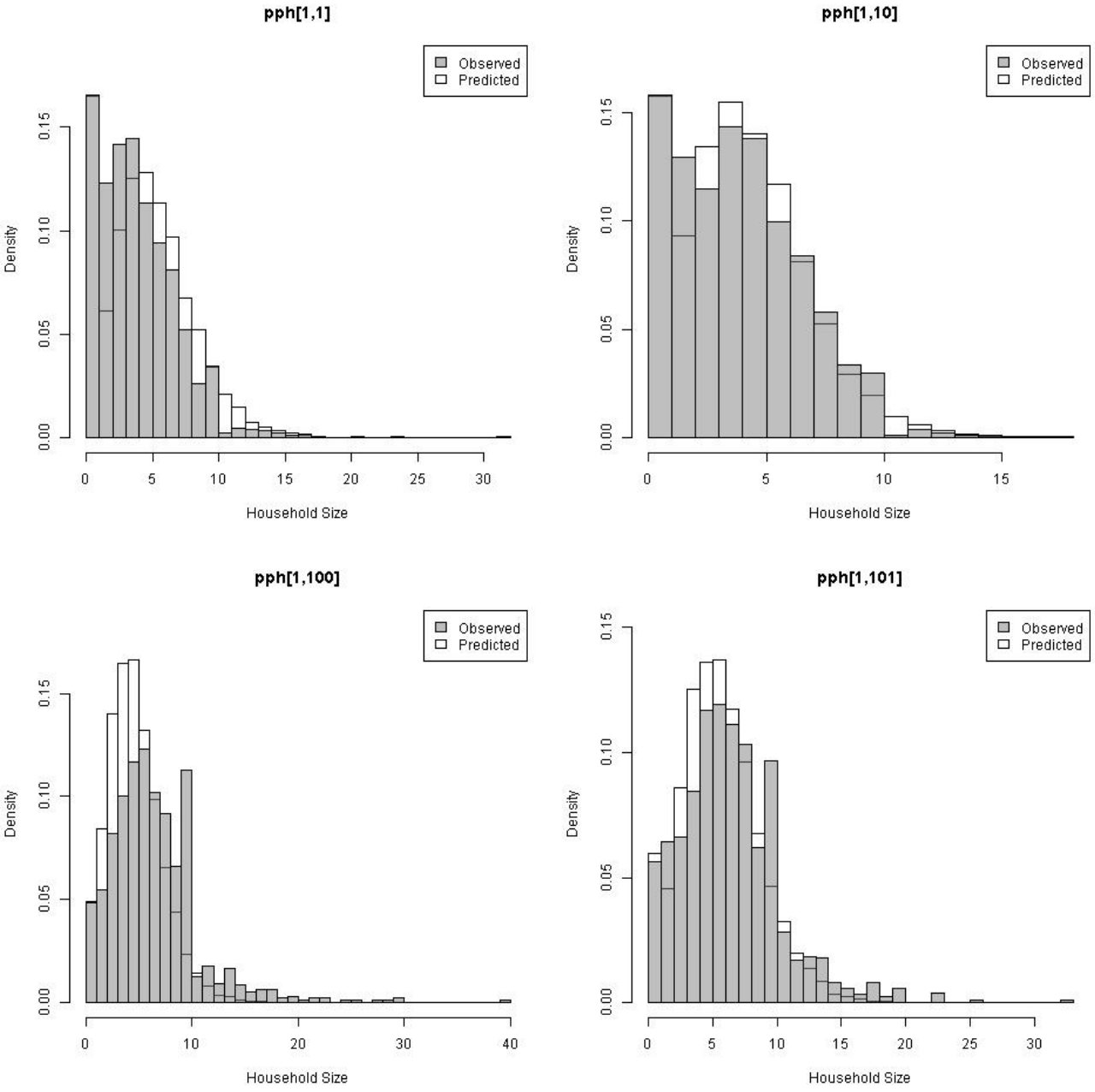

theta[2,6]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

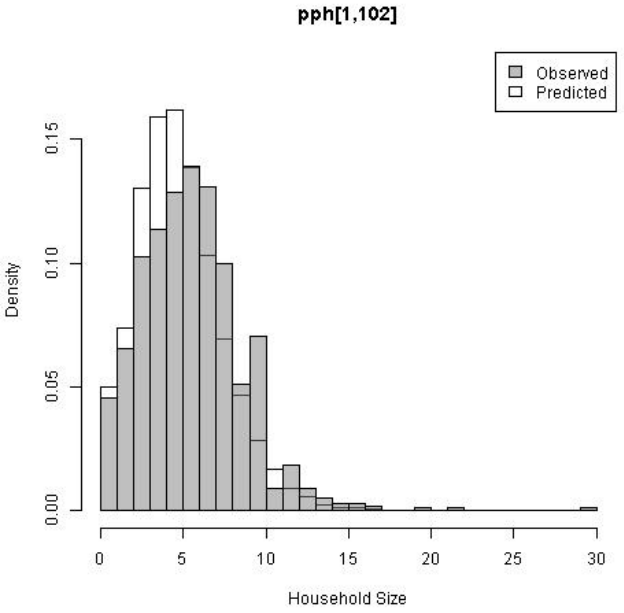

theta[2,7]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

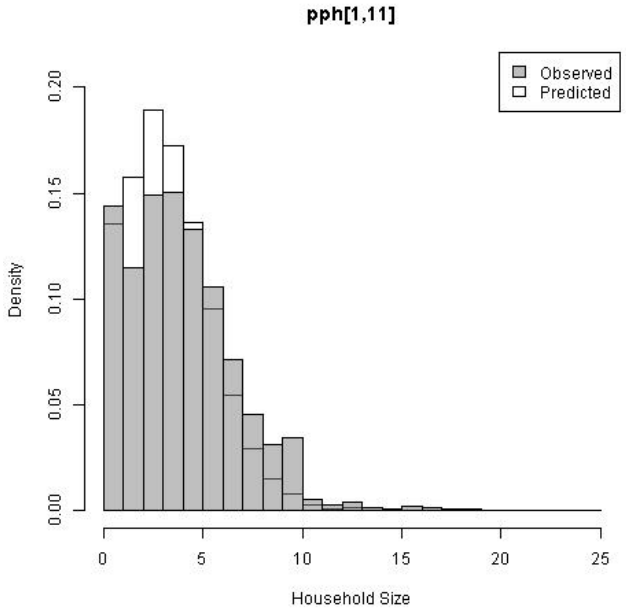

theta[2,8]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

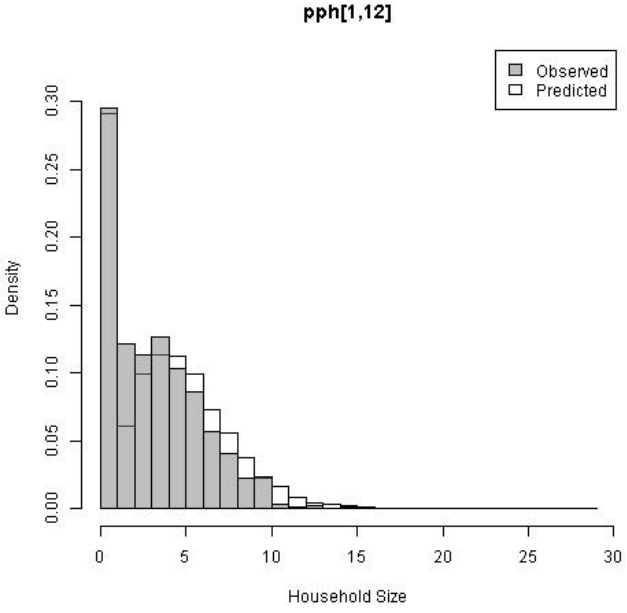

theta[2,9]
Female
Male
80+
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
1-4
<1
Age Group
10%
5%
0%
5%
10%
Proportion of Population

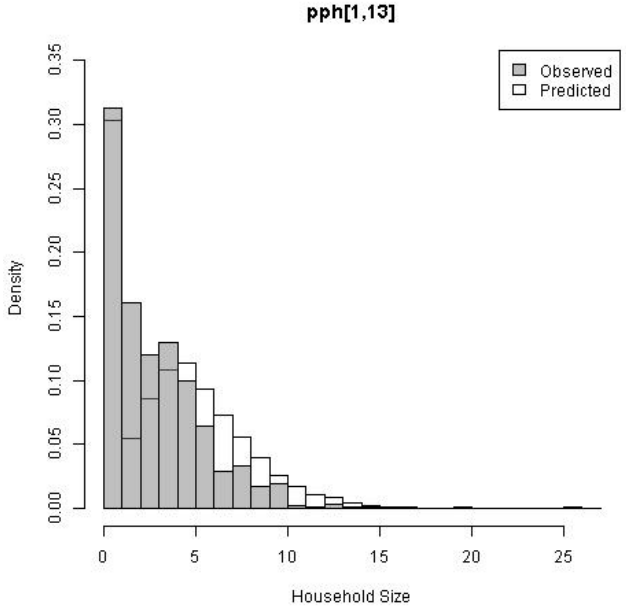
10.2 People per household

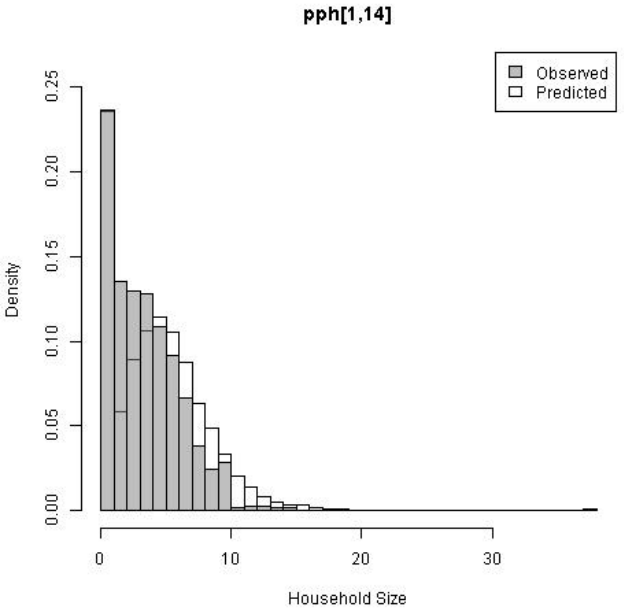


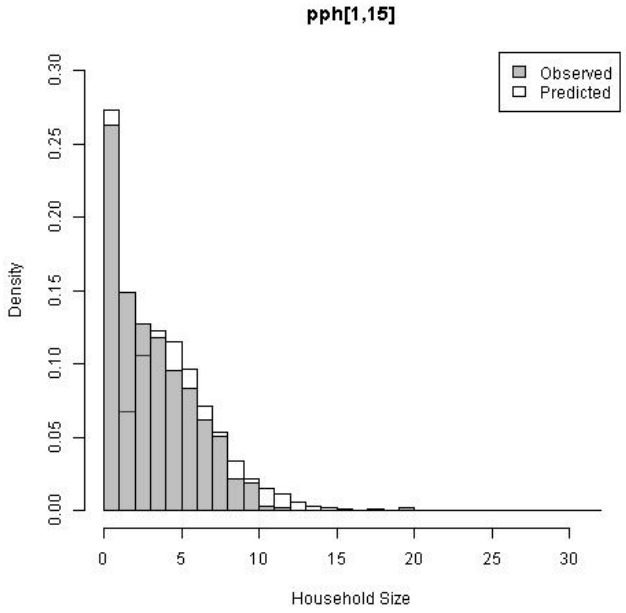

pph[1,102]
Observed
Predicted
Density
Household Size

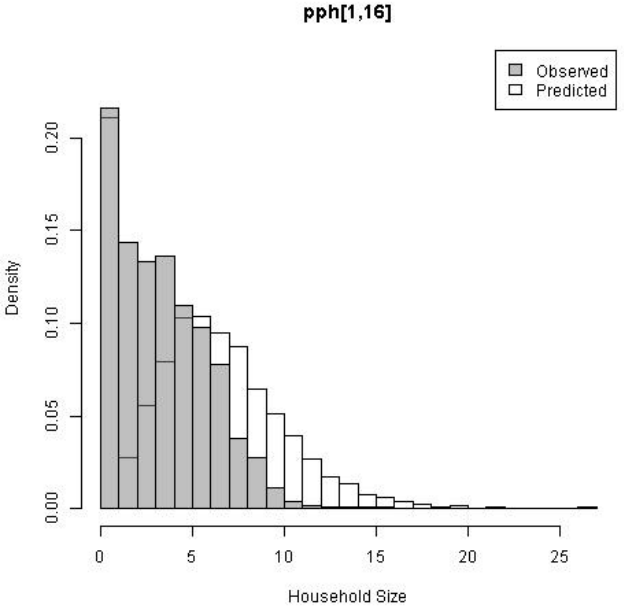

pph[1,11]
Observed
Predicted
Density
Household Size

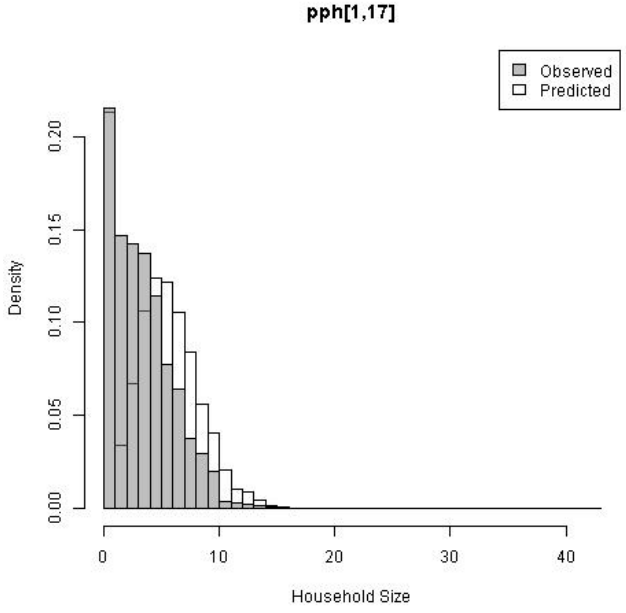

pph[1,12]
Observed
Predicted
Density
Household Size

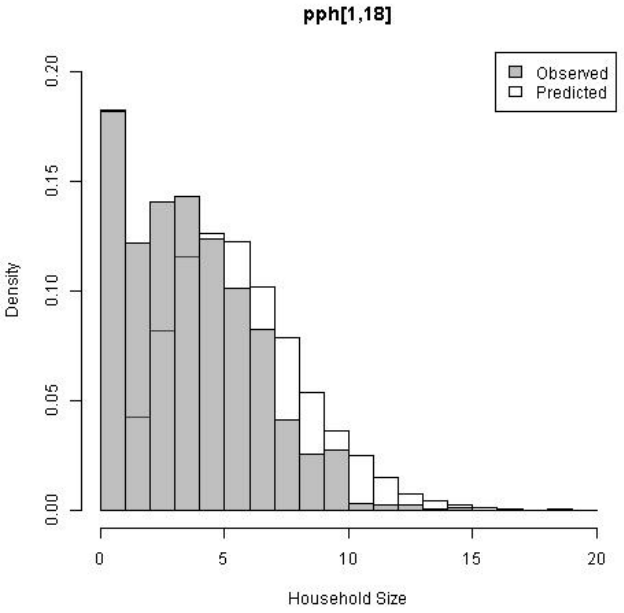

pph[1,13]
Observed
Predicted
Density
Household Size

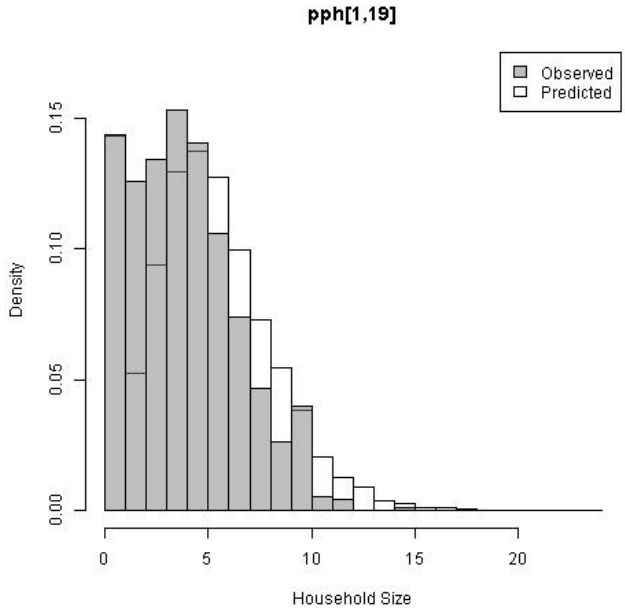

pph[1,14]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
Household Size

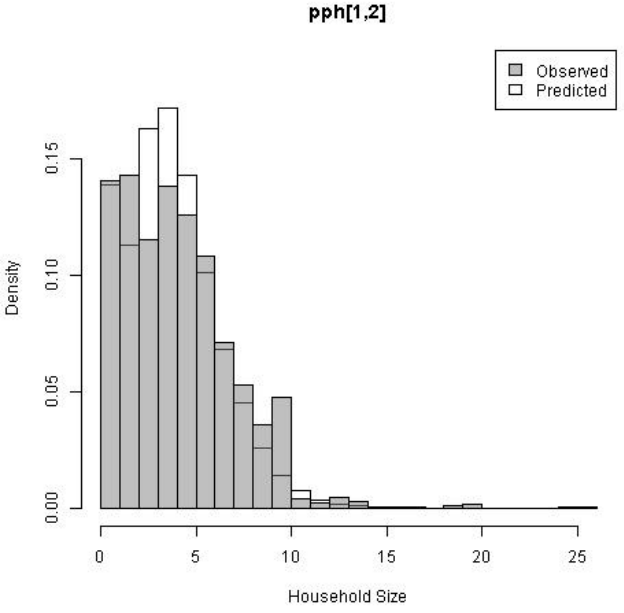

pph[1,15]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20
25
30
Household Size

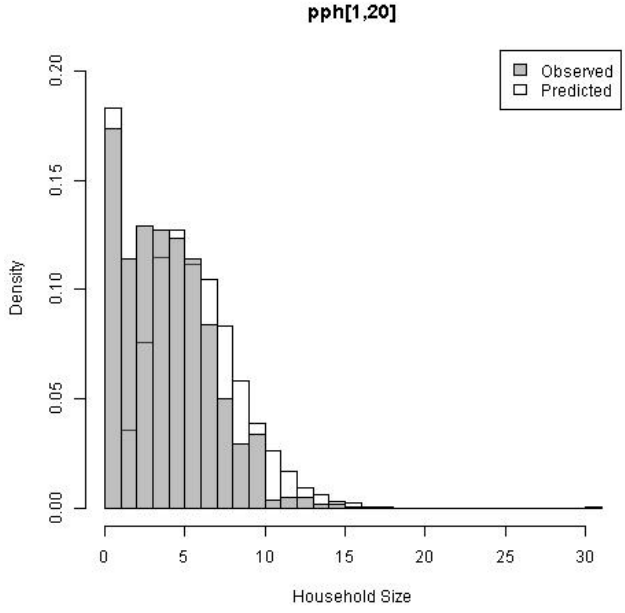

pph[1,16]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

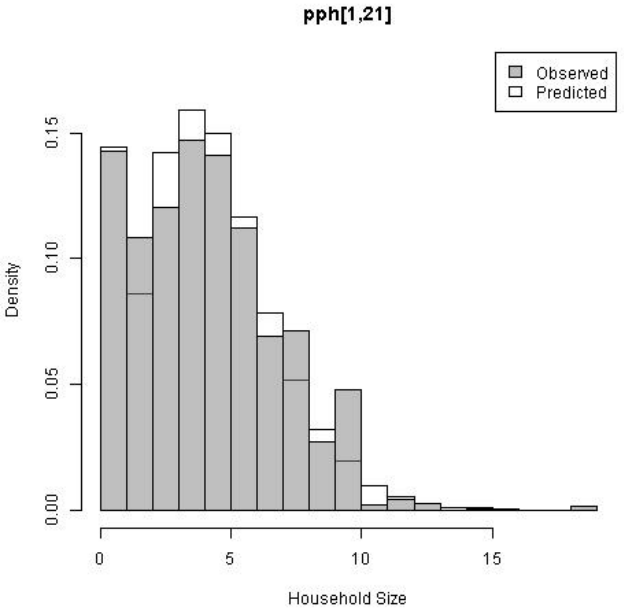

pph[1,17]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
Household Size

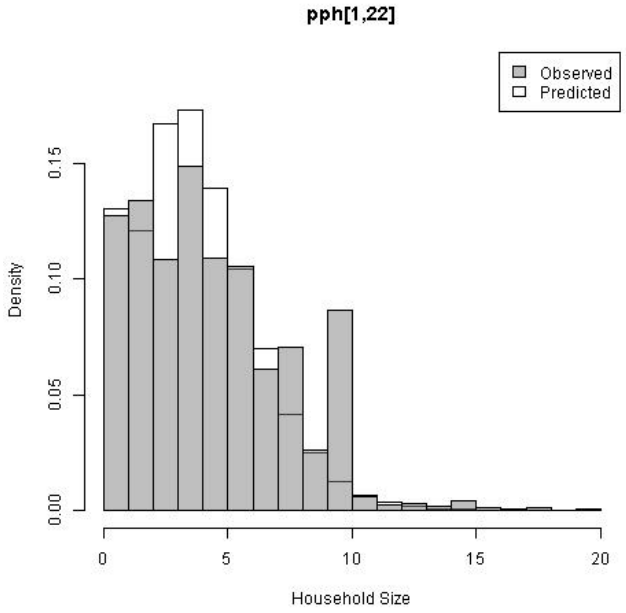

pph[1,18]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

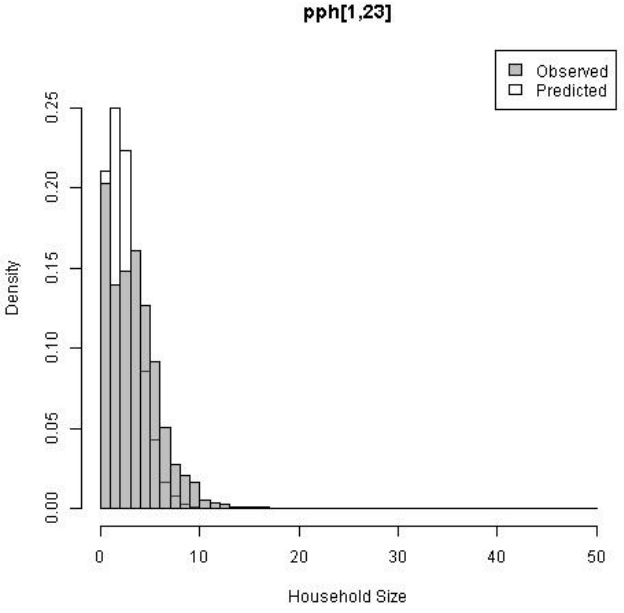

pph[1,19]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

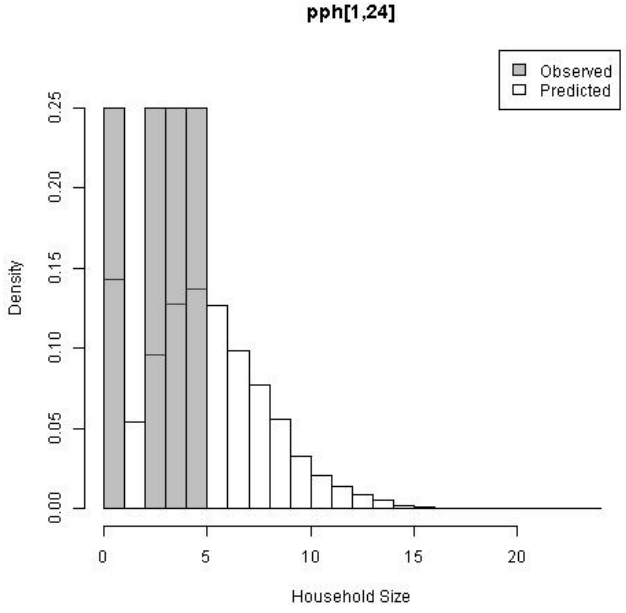

pph[1,2]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

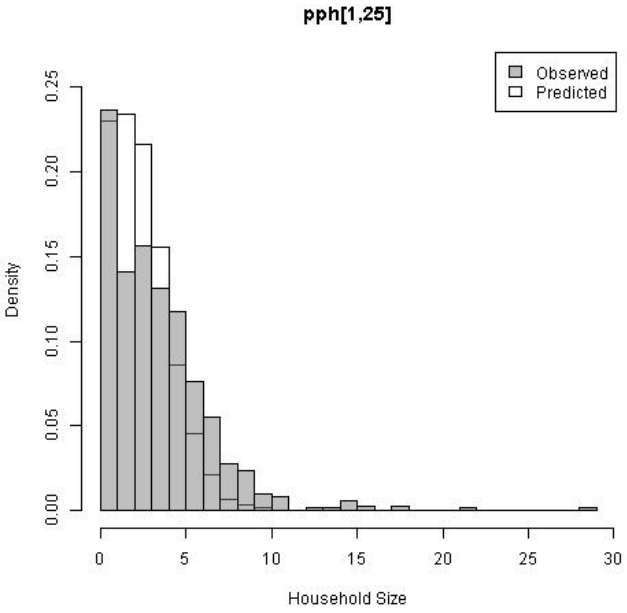

pph[1,20]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

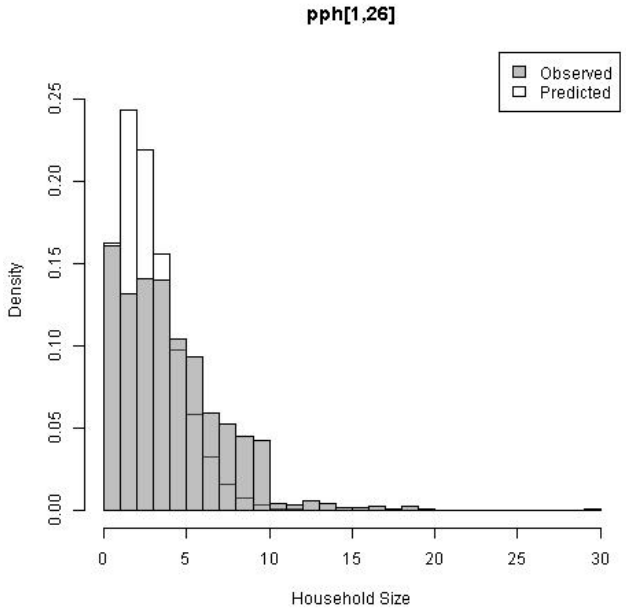

pph[1,21]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
Household Size

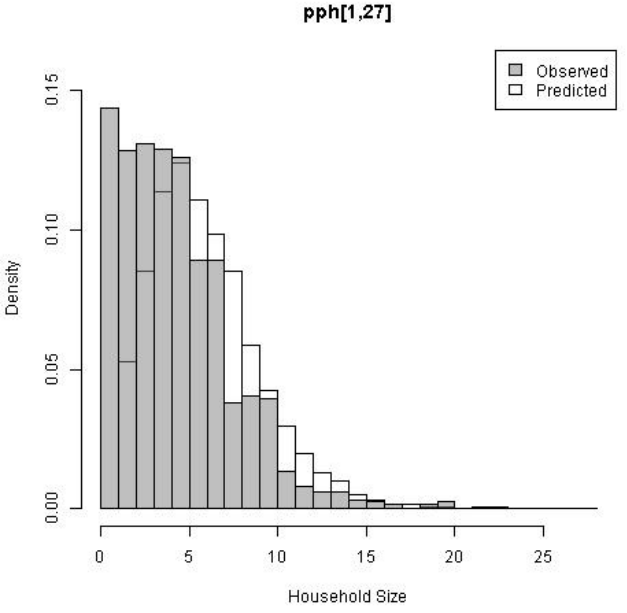

pph[1,22]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

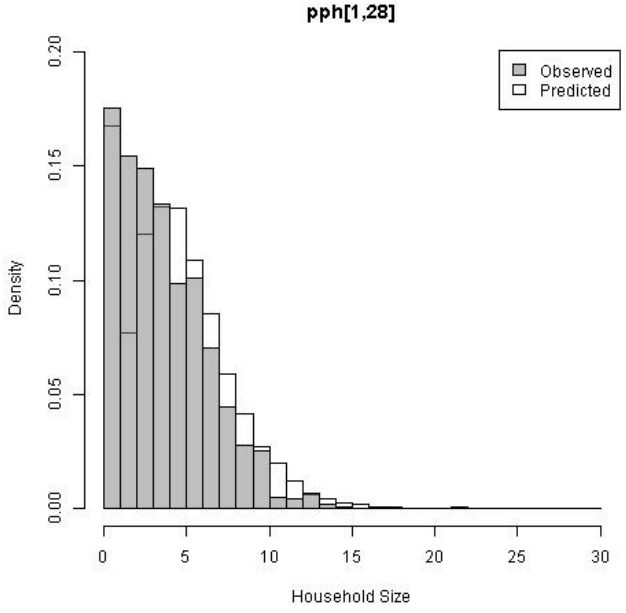

pph[1,23]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
50
Household Size

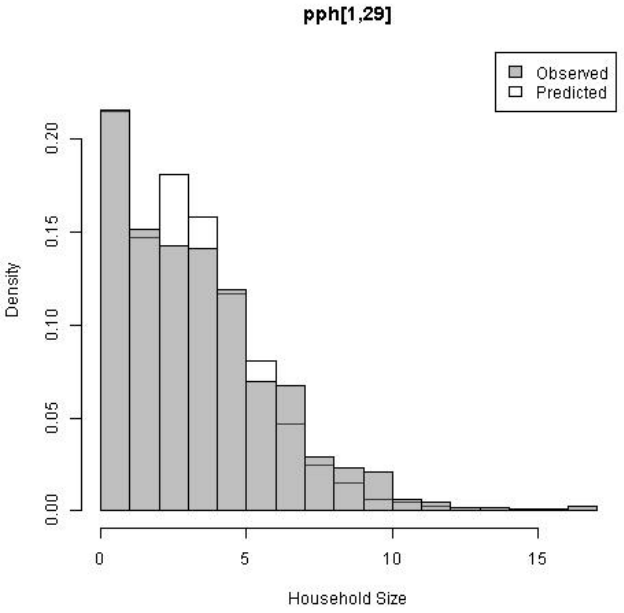

pph[1,24]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

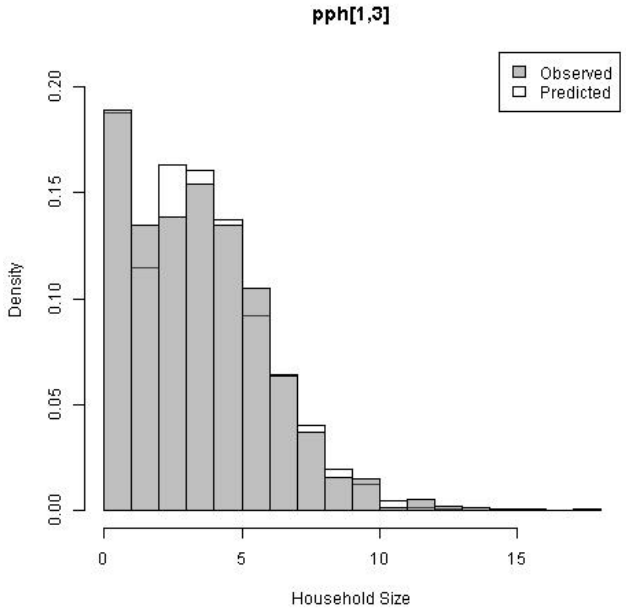

pph[1,25]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
Household Size

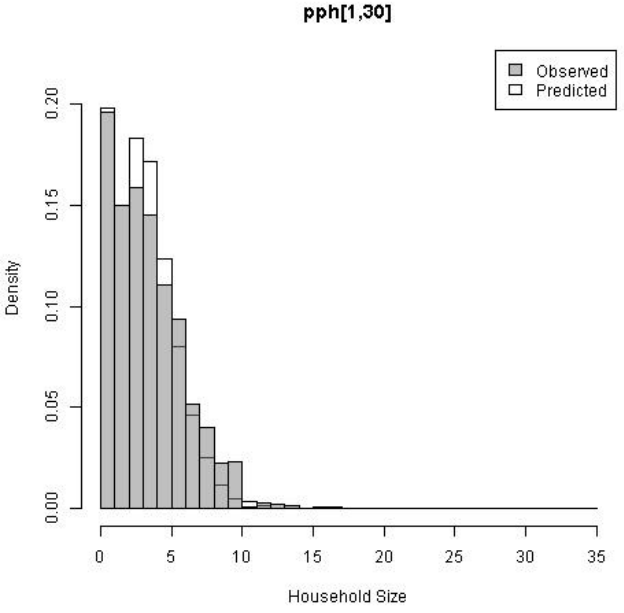

pph[1,26]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
Household Size

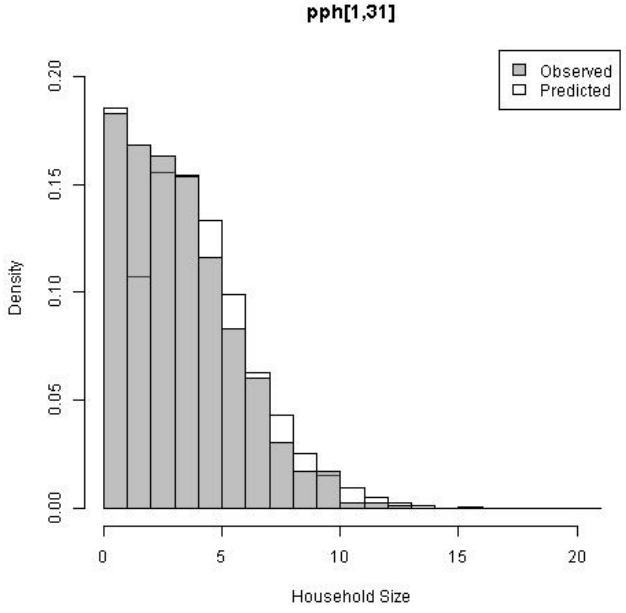

pph[1,27]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

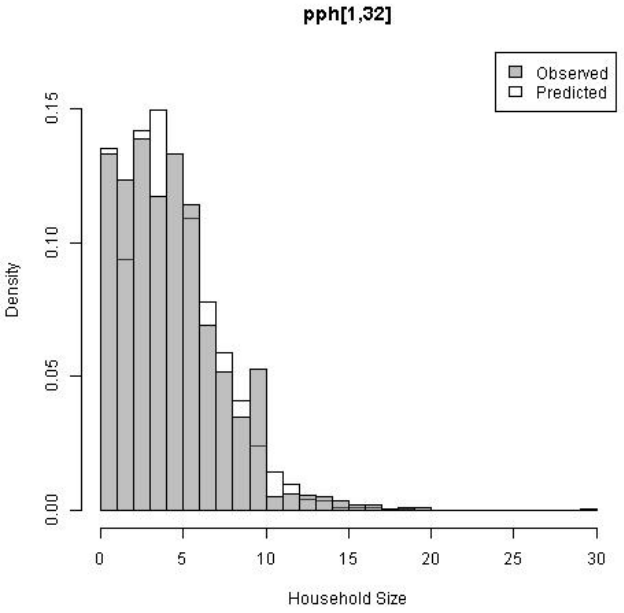

pph[1,28]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

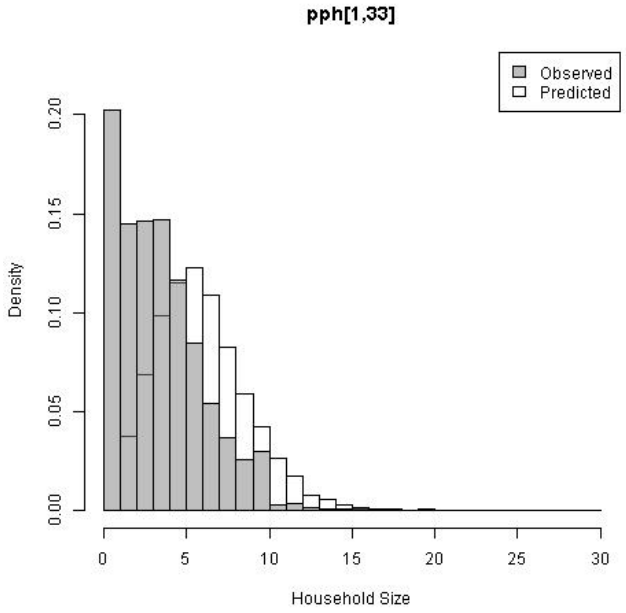

pph[1,29]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

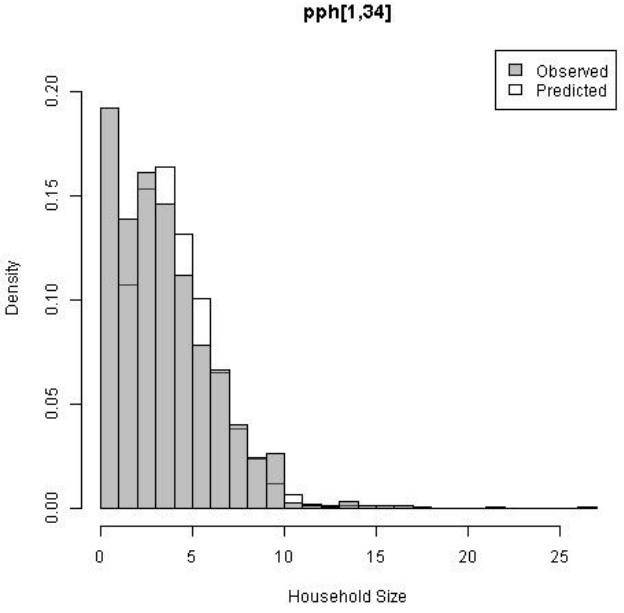

pph[1,3]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

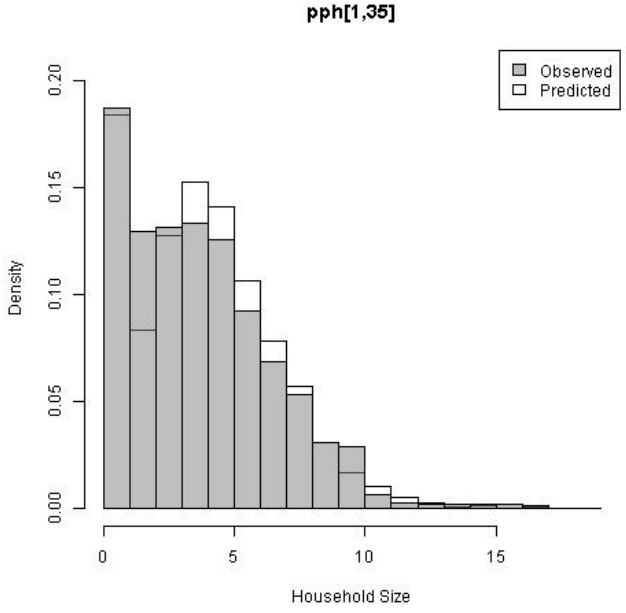

pph[1,30]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
35
Household Size

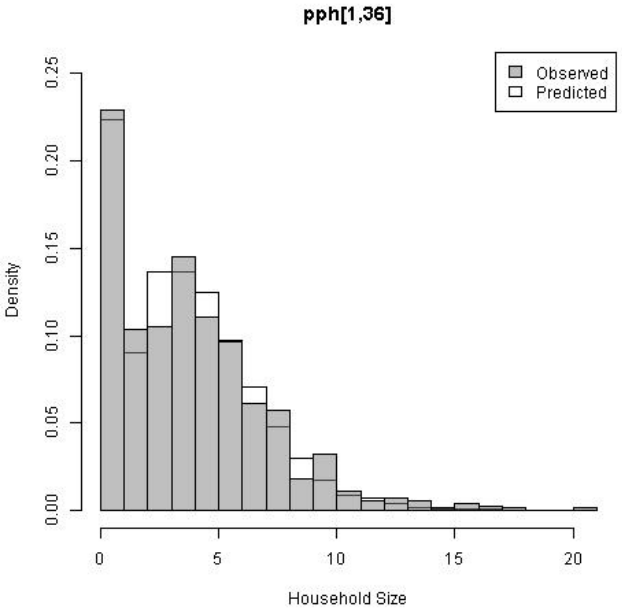

pph[1,31]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

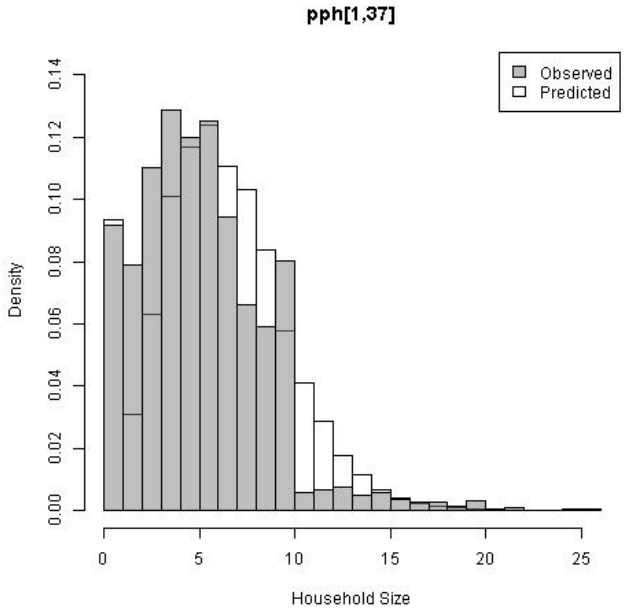

pph[1,32]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

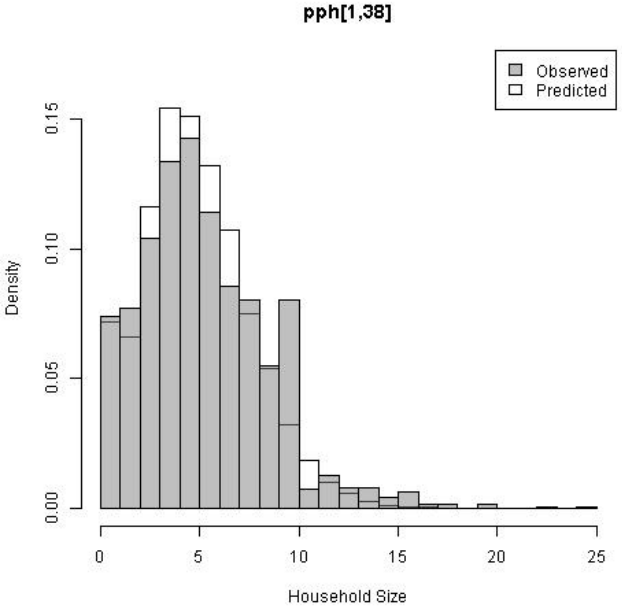

pph[1,33]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

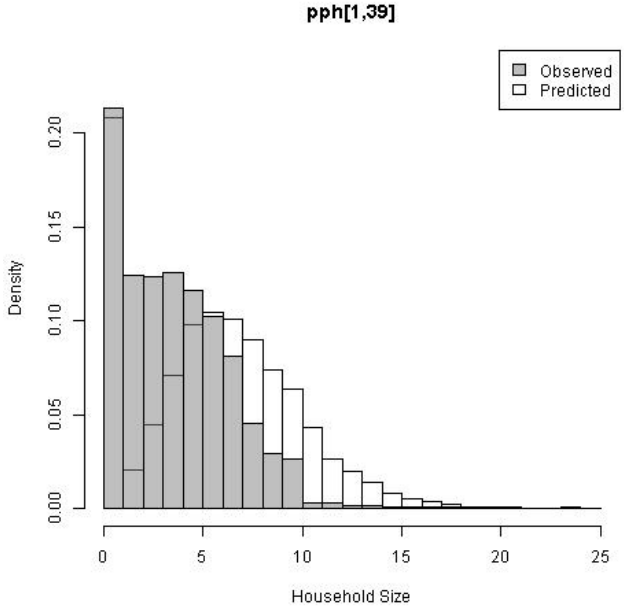

pph[1,34]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

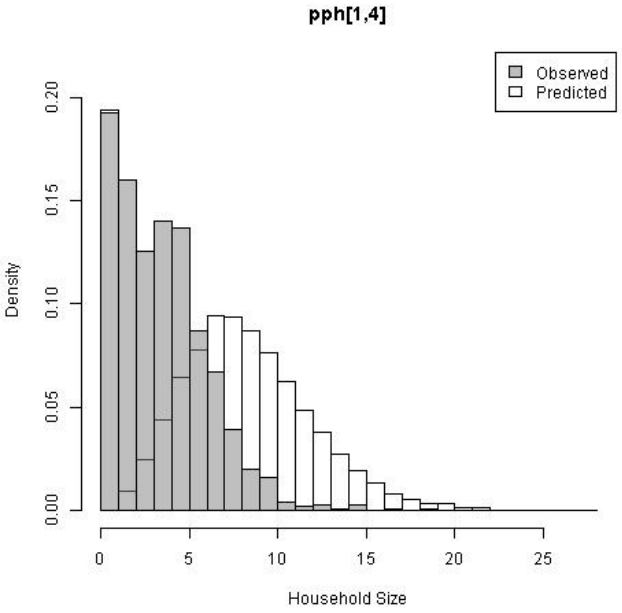

pph[1,35]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

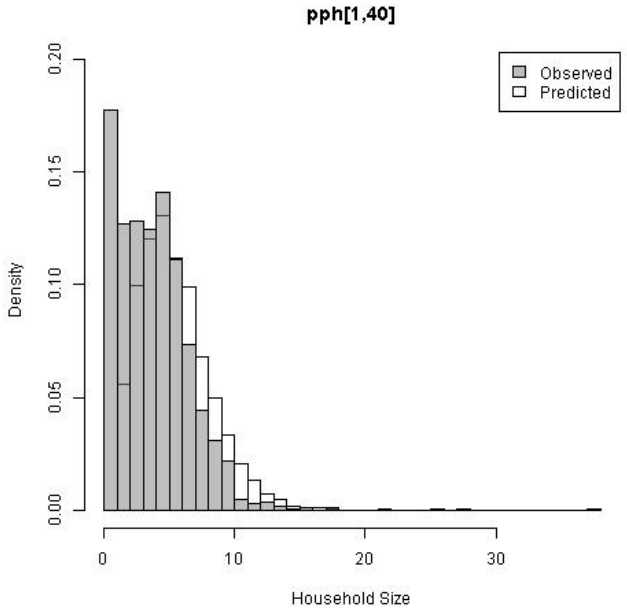

pph[1,36]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

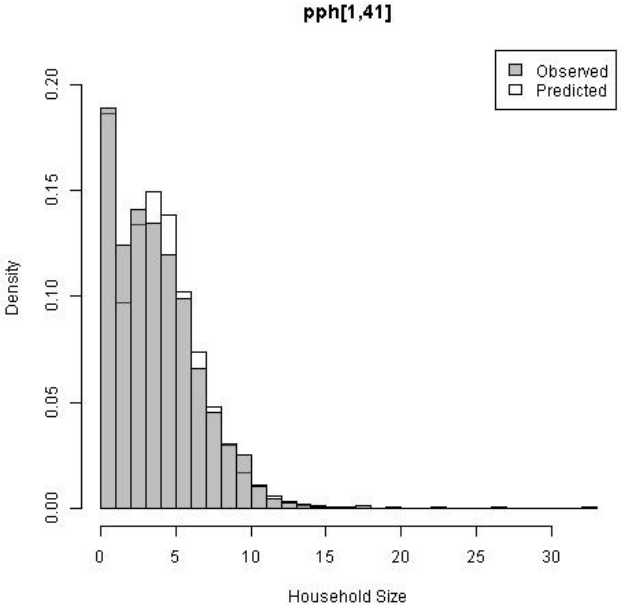

pph[1,37]
Observed
Predicted
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
5
10
15
20
25
Household Size

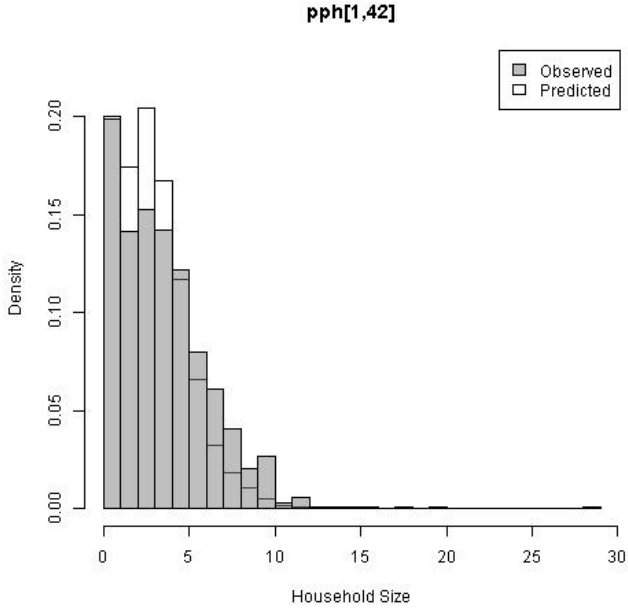

pph[1,38]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

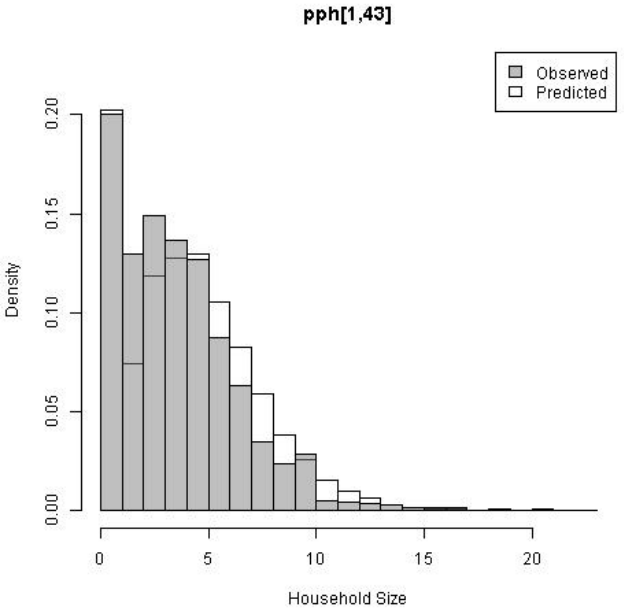

pph[1,39]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

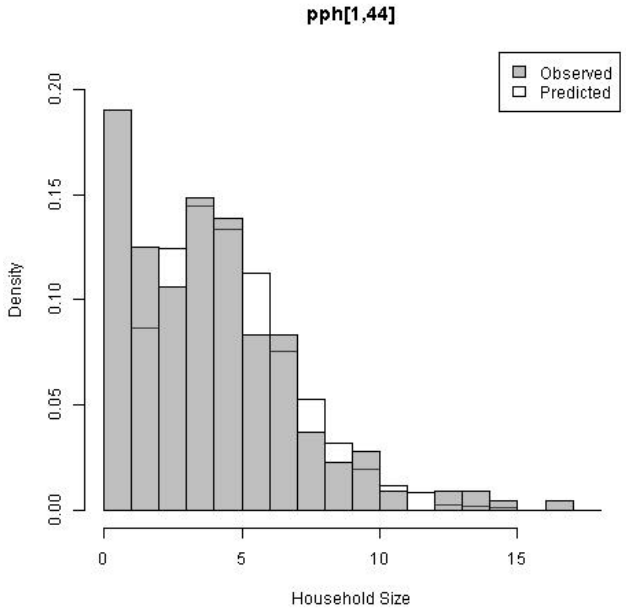

pph[1,4]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

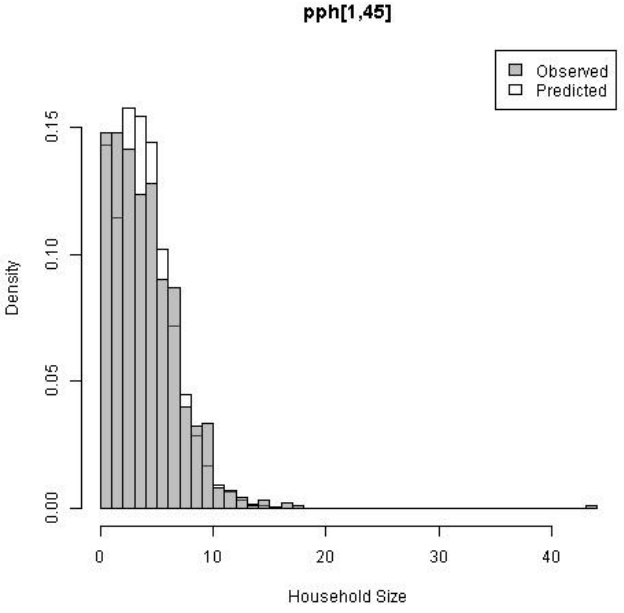

pph[1,40]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
Household Size

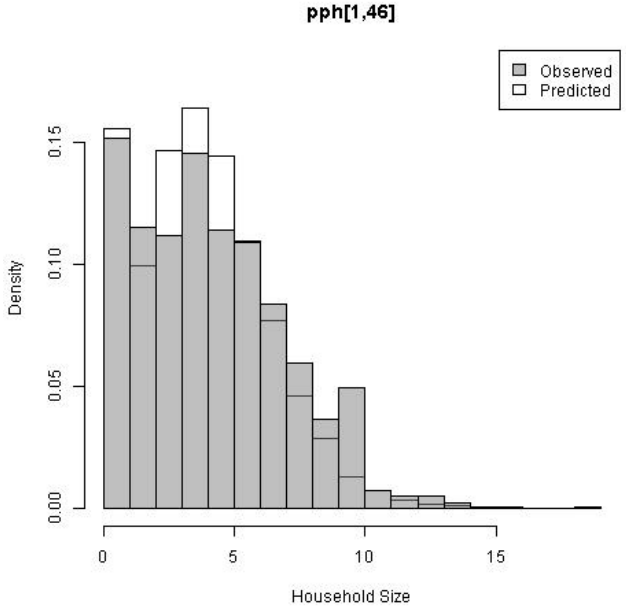

pph[1,41]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

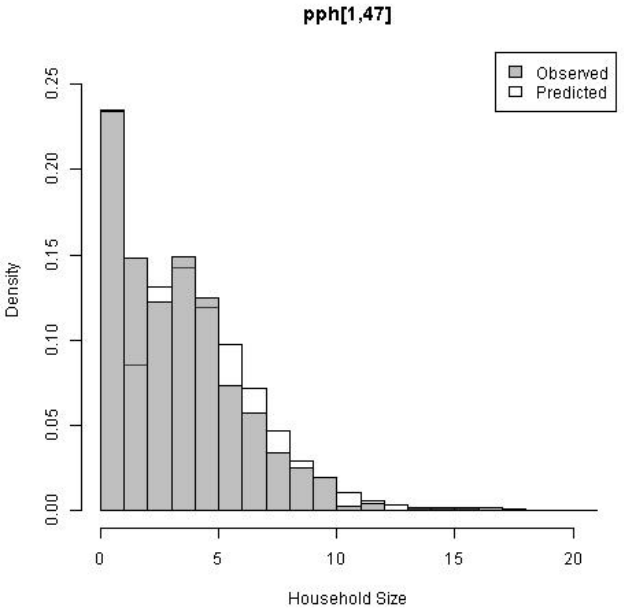

pph[1,42]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

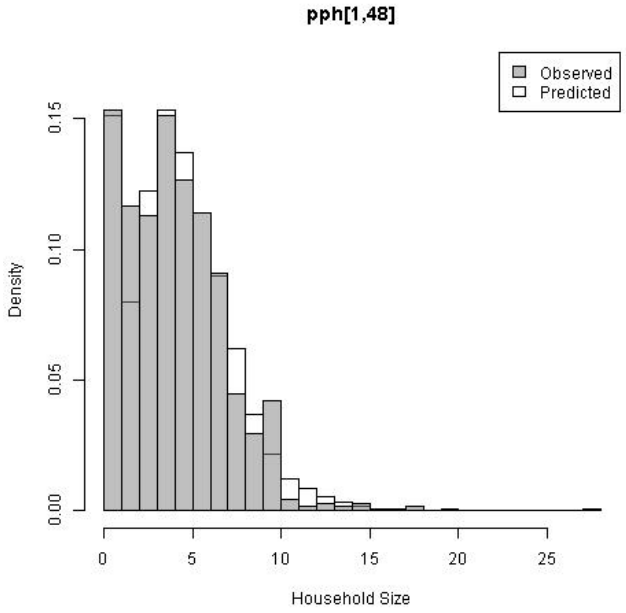

pph[1,43]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

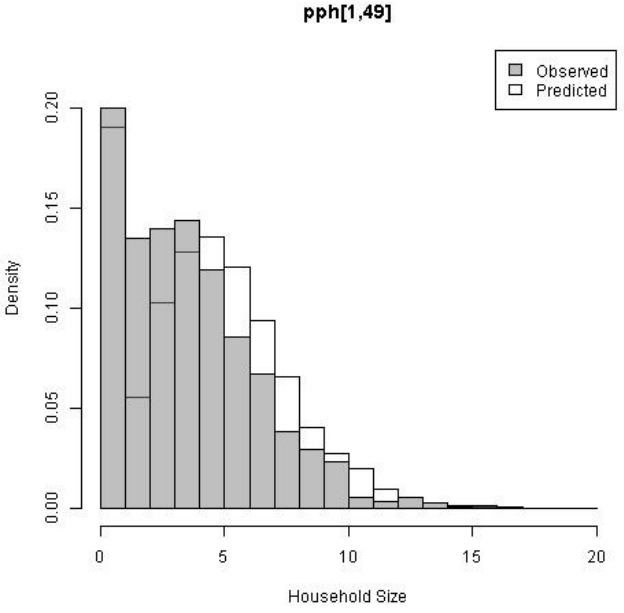

pph[1,44]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

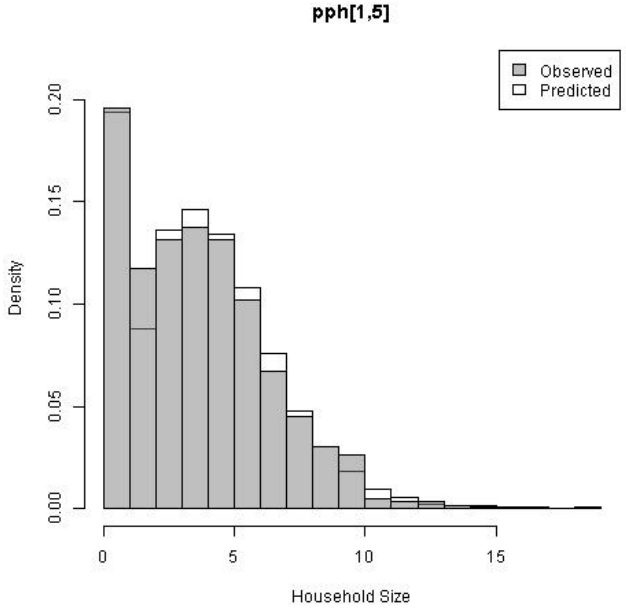

pph[1,45]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size

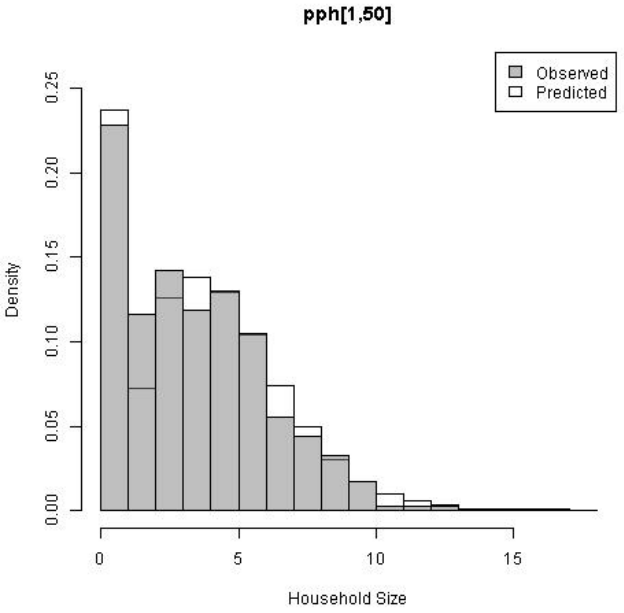

pph[1,46]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
Household Size

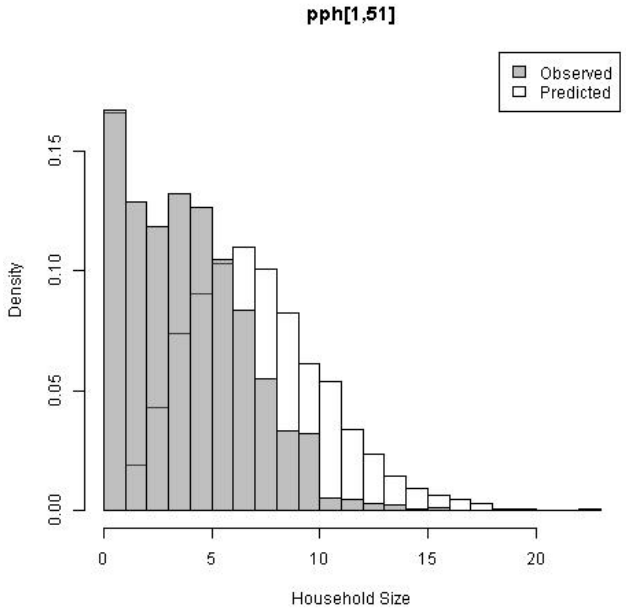

pph[1,47]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

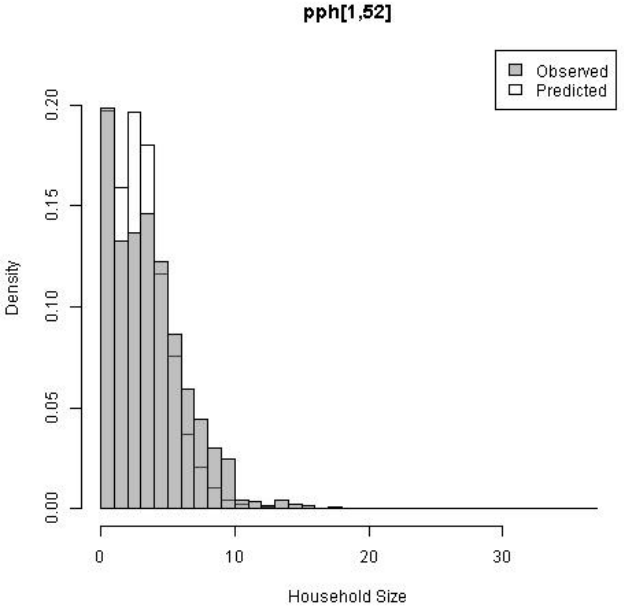

pph[1,48]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

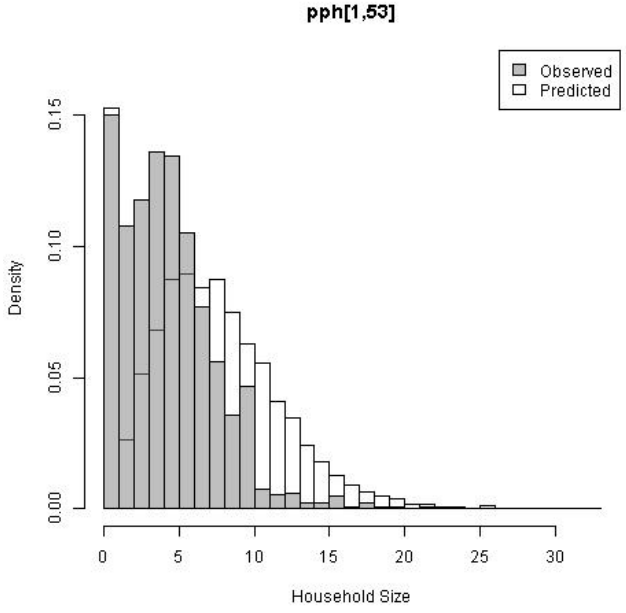

pph[1,49]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

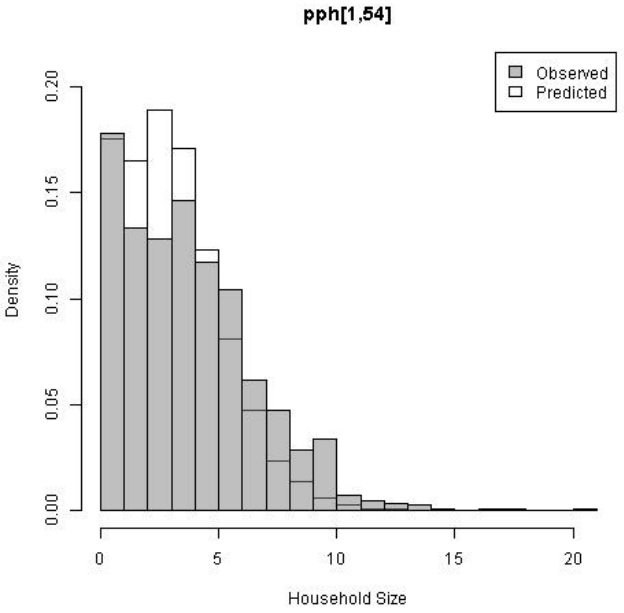

pph[1,5]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

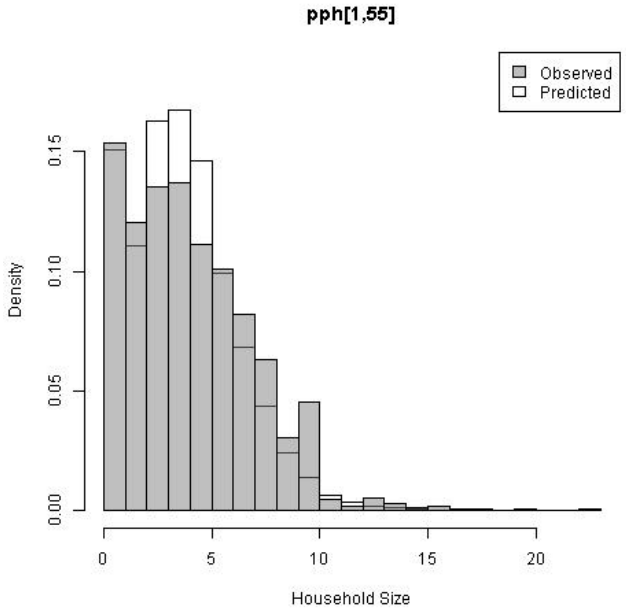

pph[1,50]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
Household Size

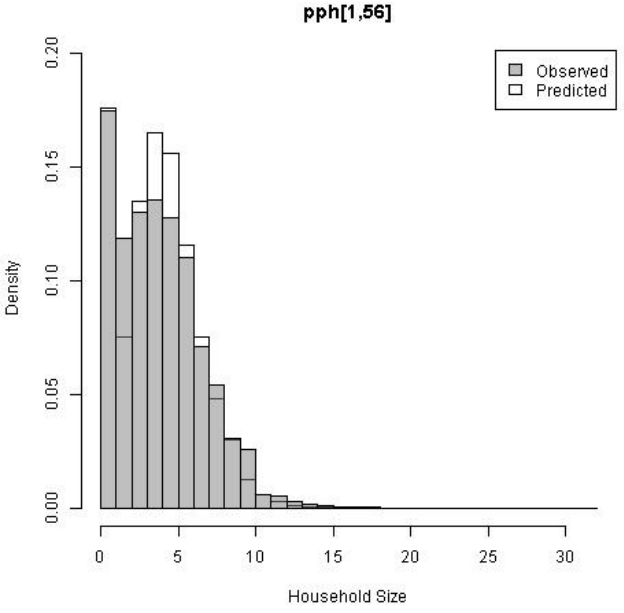

pph[1,51]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

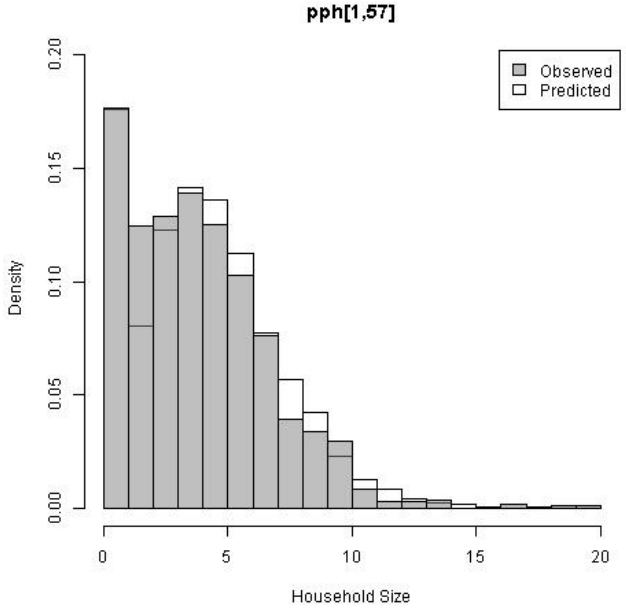

pph[1,52]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
Household Size

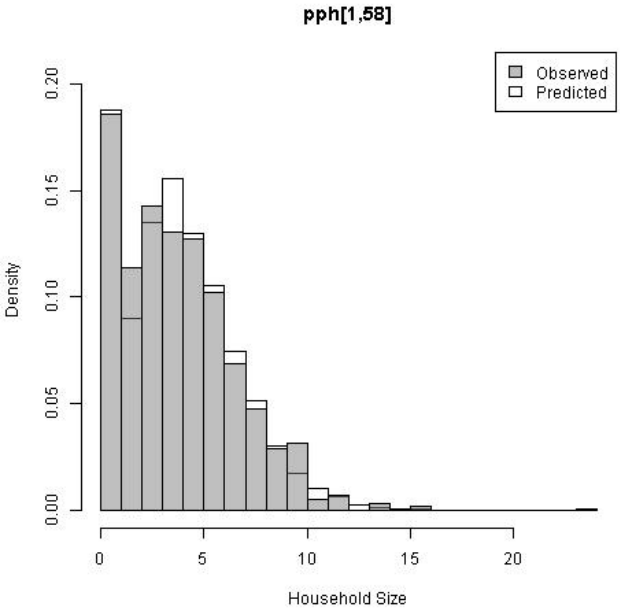

pph[1,53]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

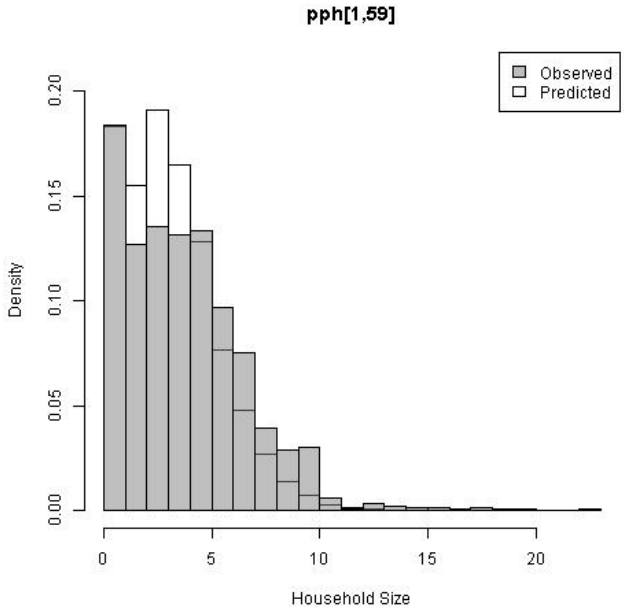

pph[1,54]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

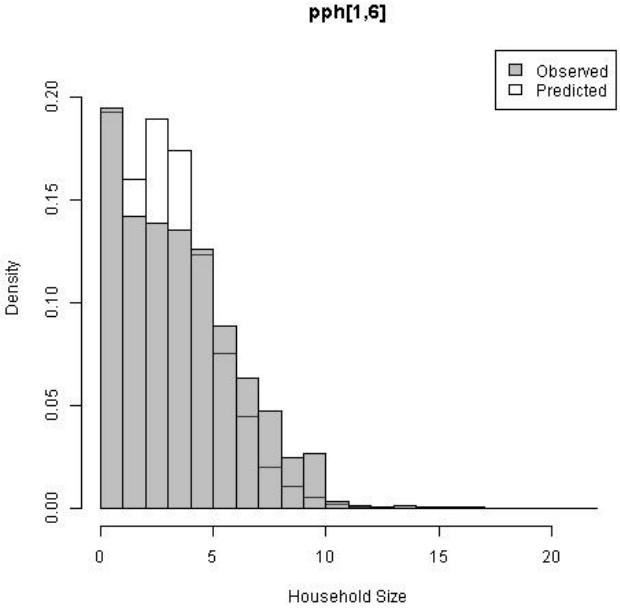

pph[1,55]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

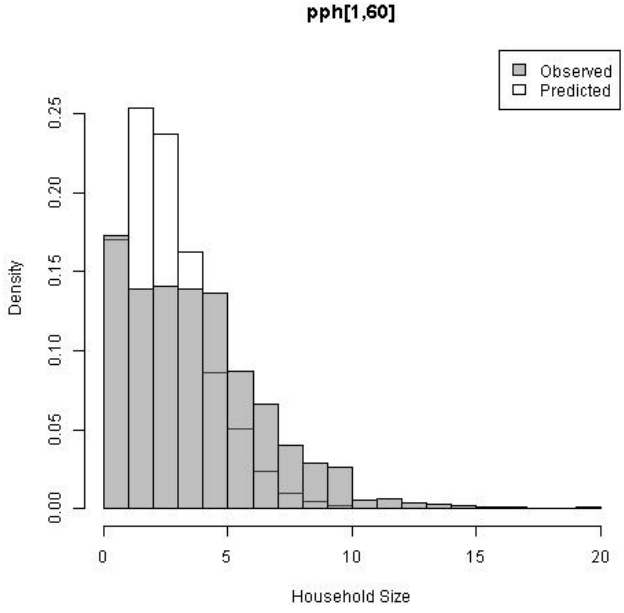

pph[1,56]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

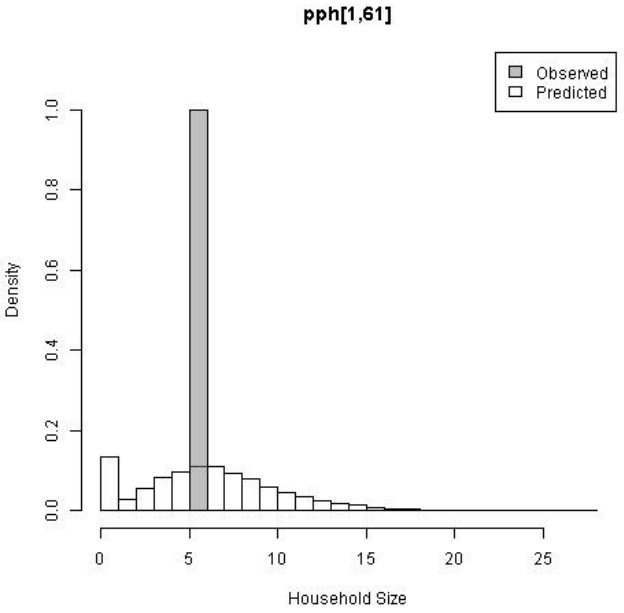

pph[1,57]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

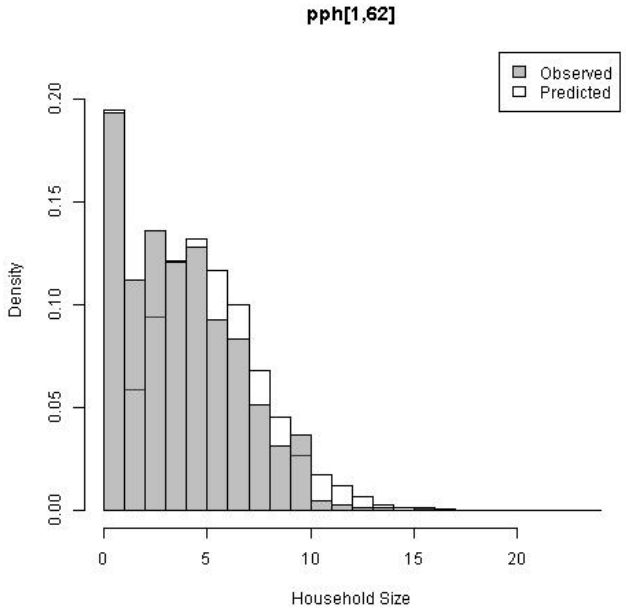

pph[1,58]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

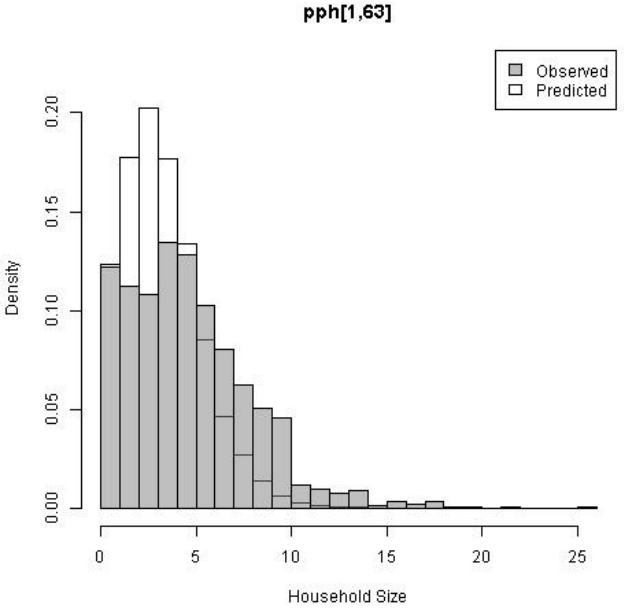

pph[1,59]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

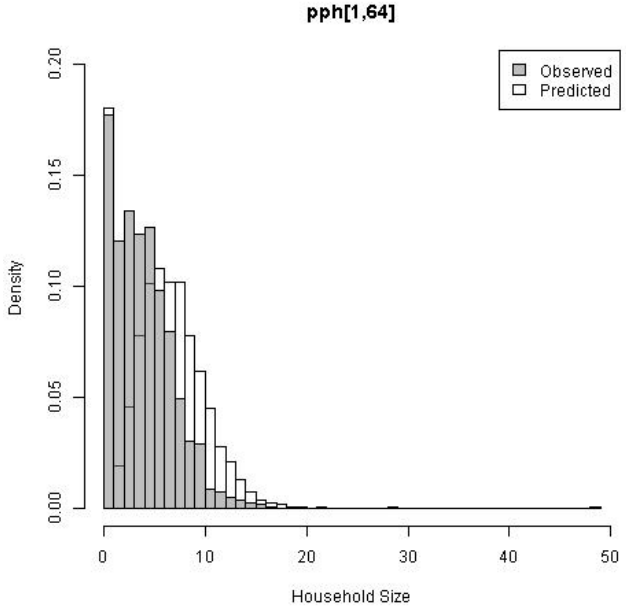

pph[1,6]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

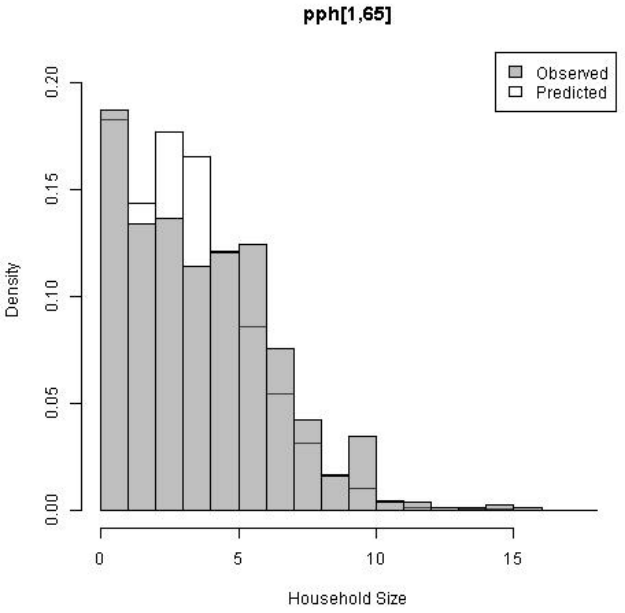

pph[1,60]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

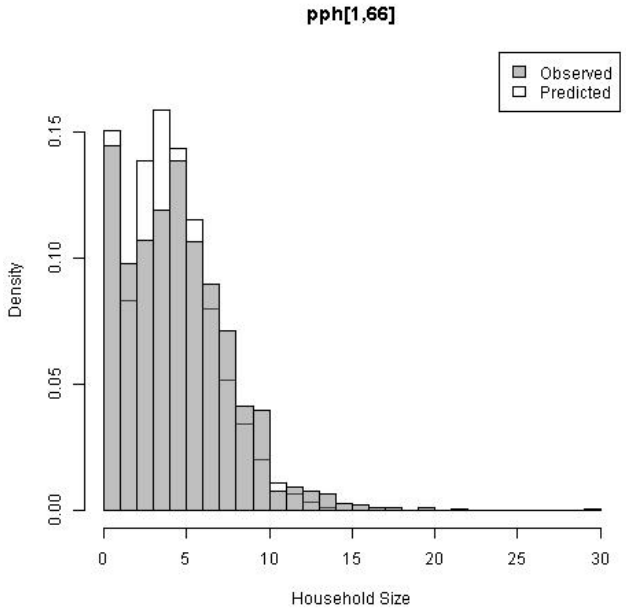

pph[1,61]
Observed
Predicted
Density
Household Size

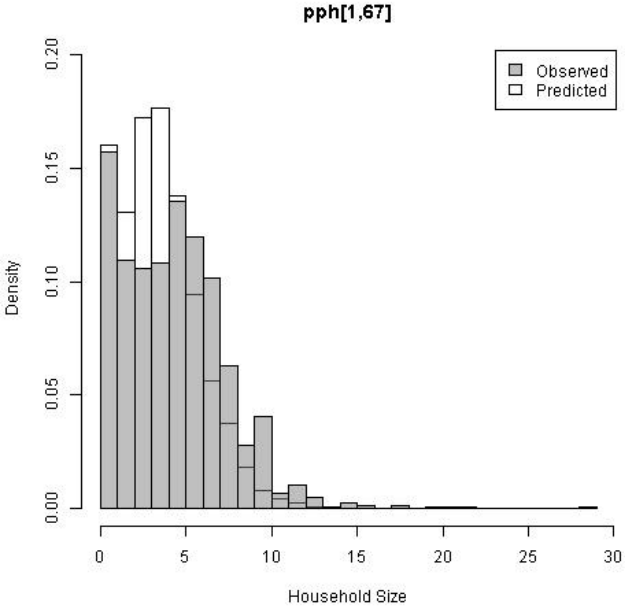

pph[1,62]
Observed
Predicted
Density
Household Size

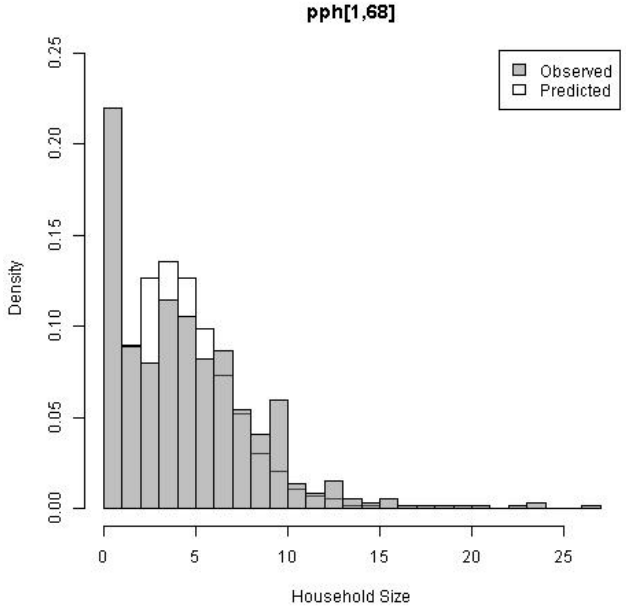

pph[1,63]
Observed
Predicted
Density
Household Size

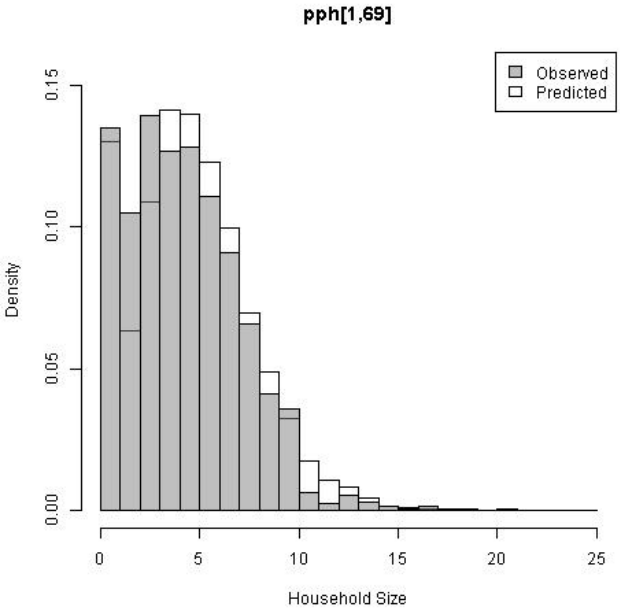

pph[1,64]
Observed
Predicted
Density
Household Size

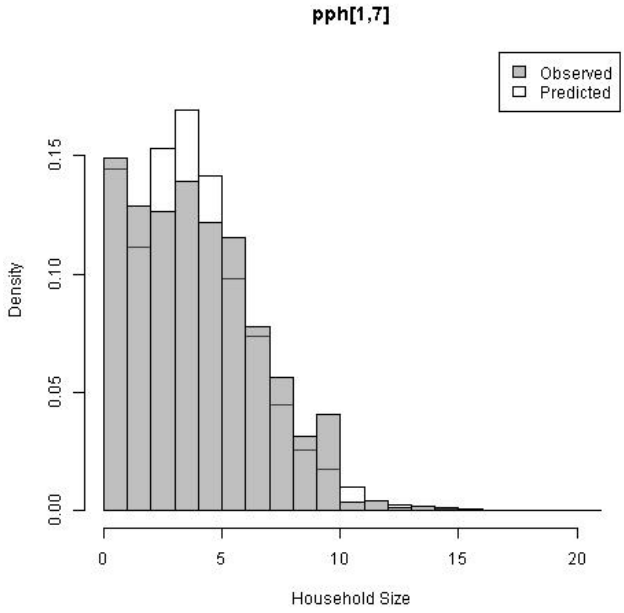

pph[1,65]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

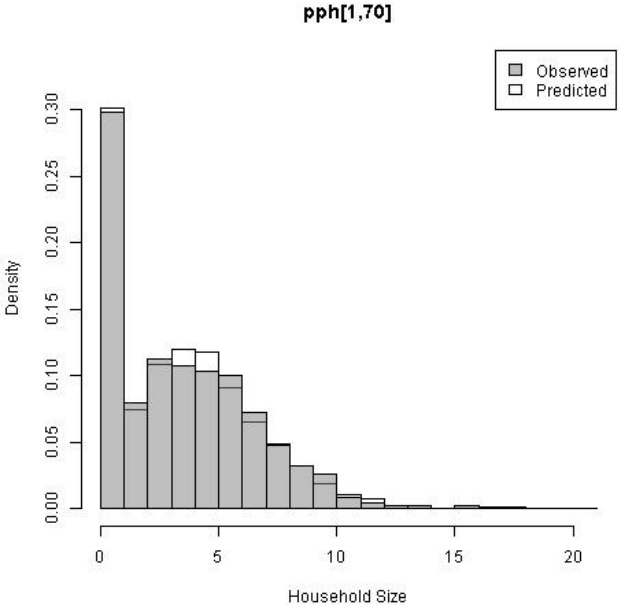

pph[1,66]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

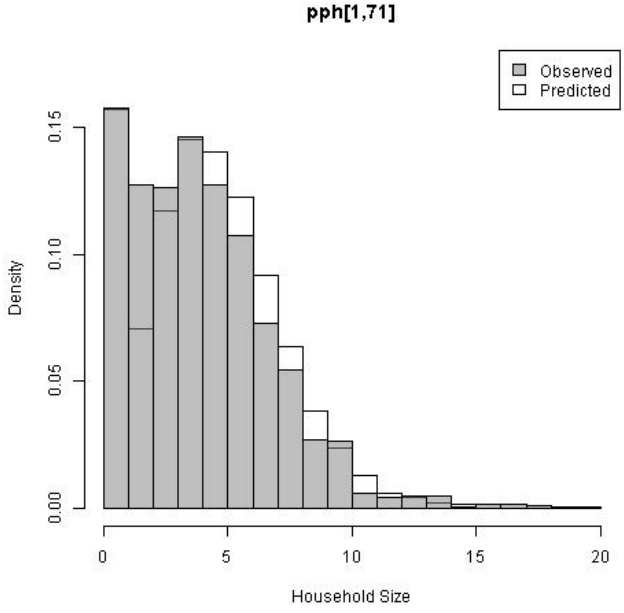

pph[1,67]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

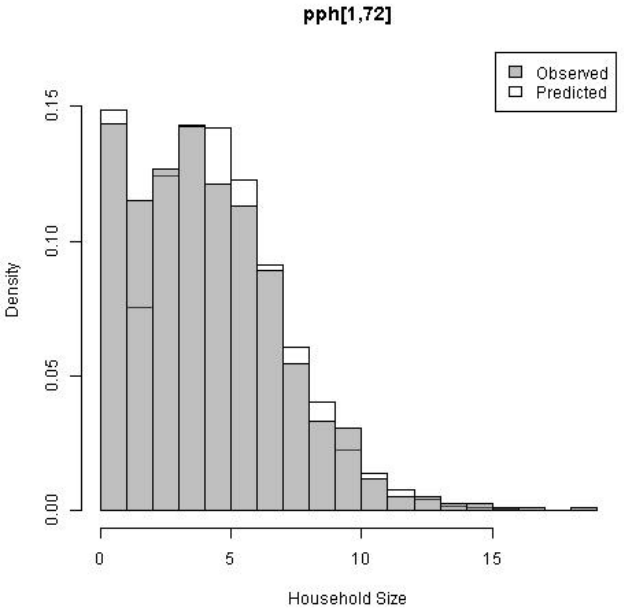

pph[1,68]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
Household Size

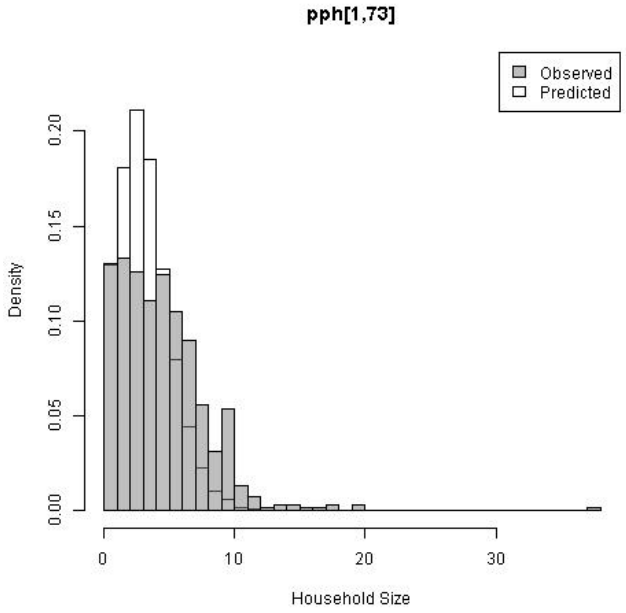

pph[1,69]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

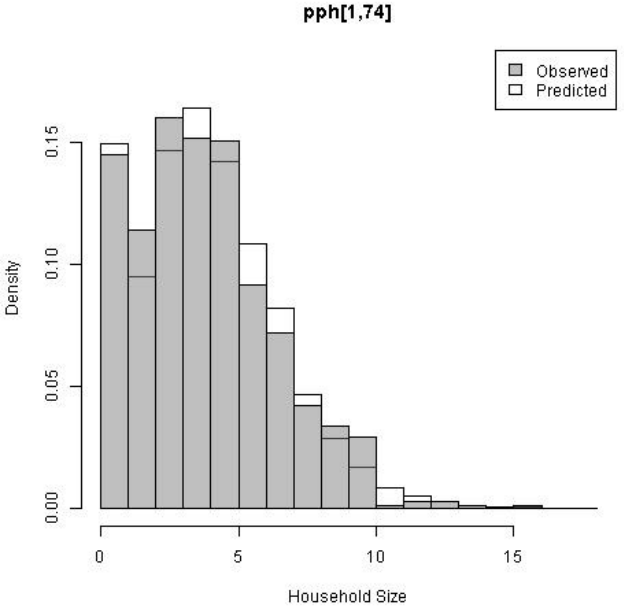

pph[1,7]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

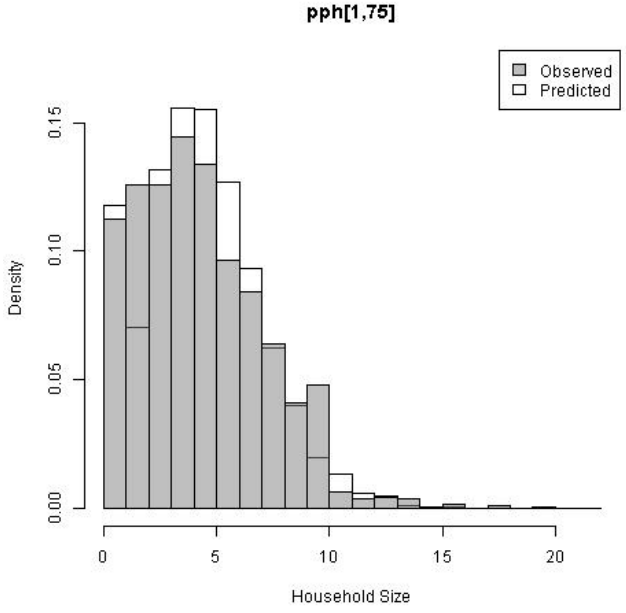

pph[1,70]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20
Household Size

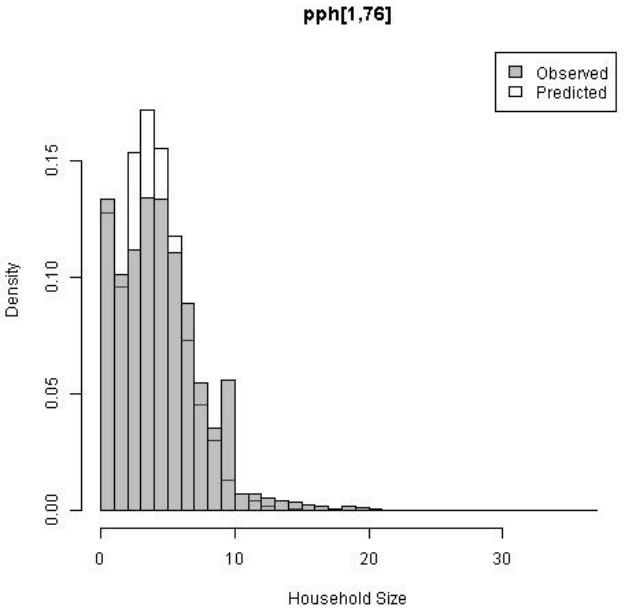

pph[1,71]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

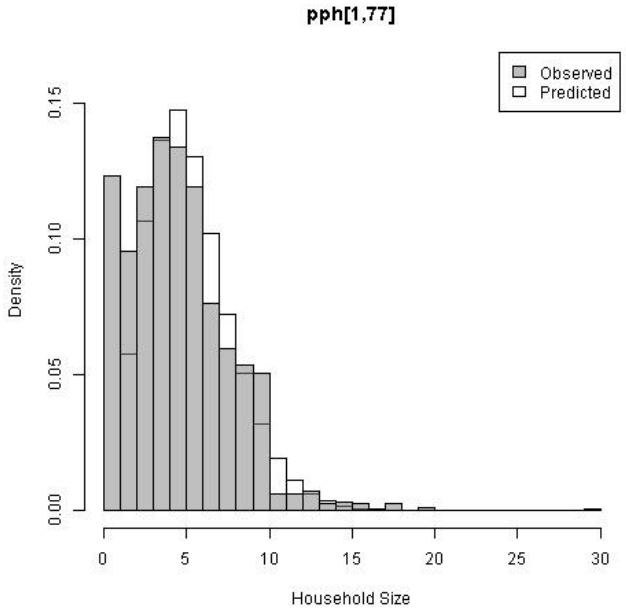

pph[1,72]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
Household Size

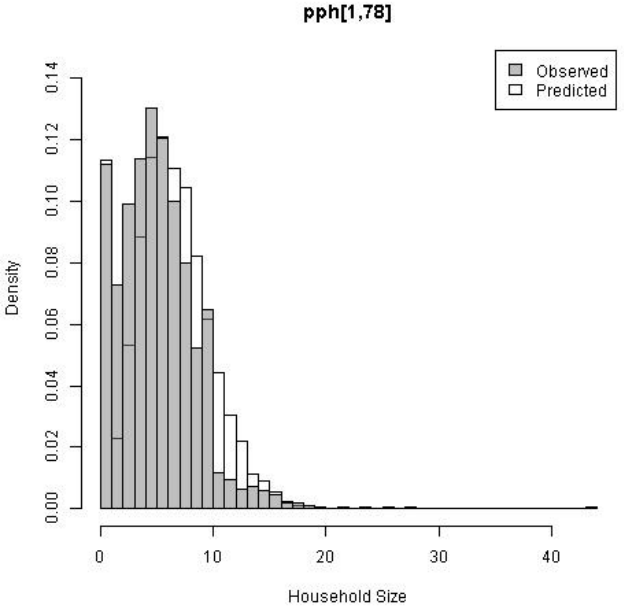

pph[1,73]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
Household Size

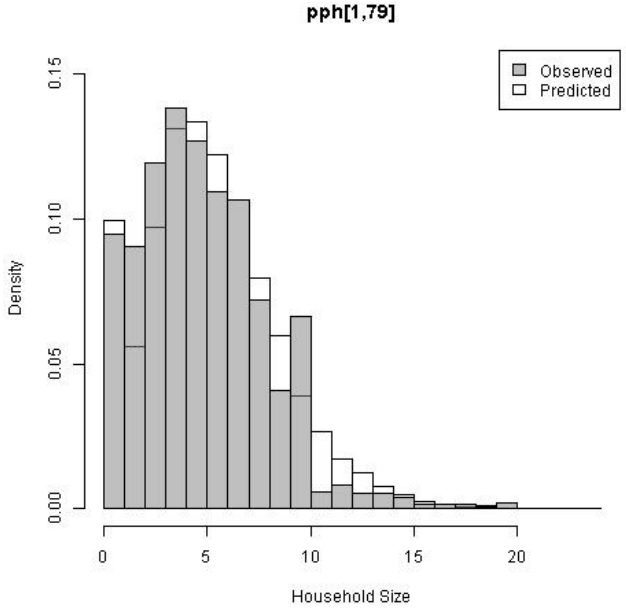

pph[1,74]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
Household Size

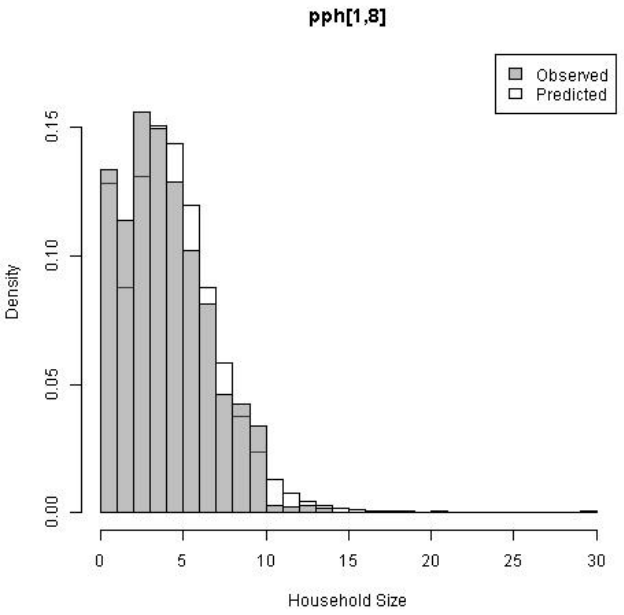

pph[1,75]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

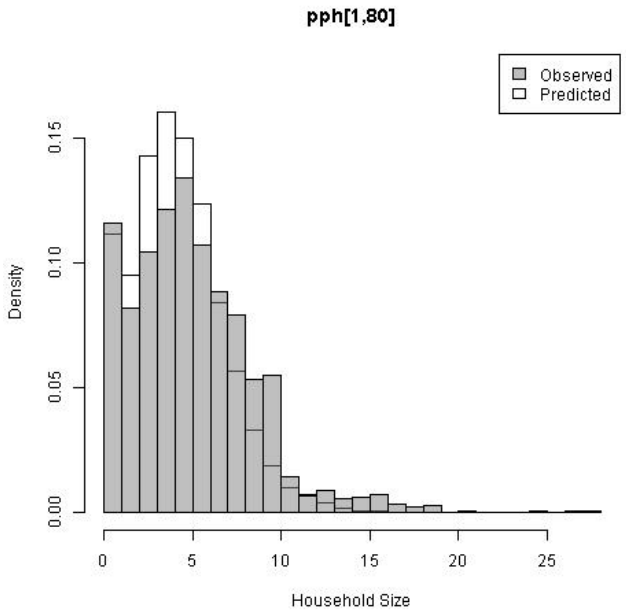

pph[1,76]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
Household Size

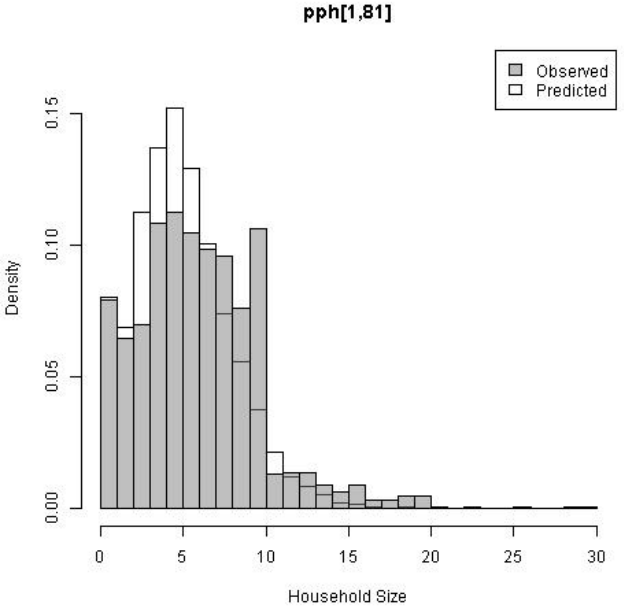

pph[1,77]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

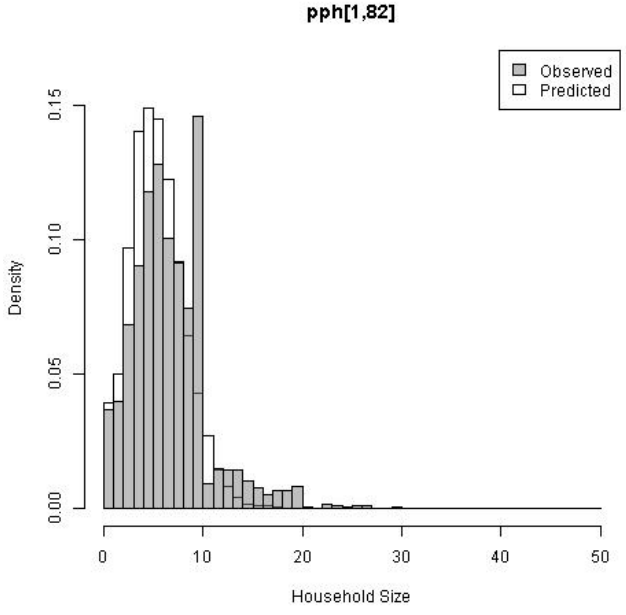

pph[1,78]
Observed
Predicted
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
10
20
30
40
Household Size

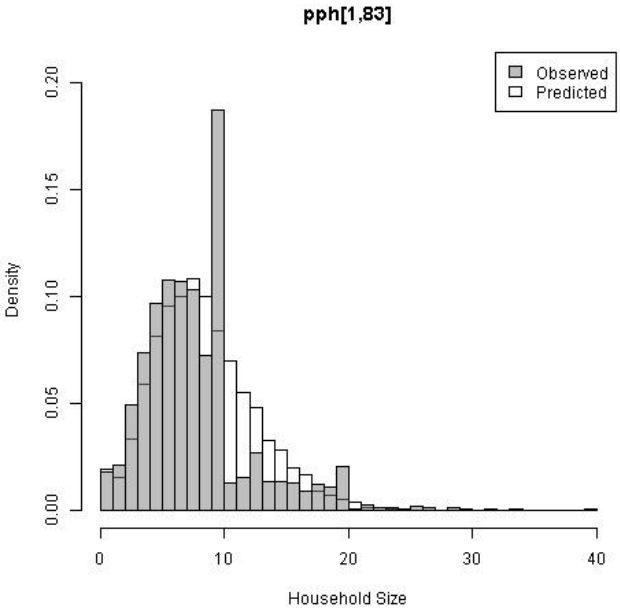

pph[1,79]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

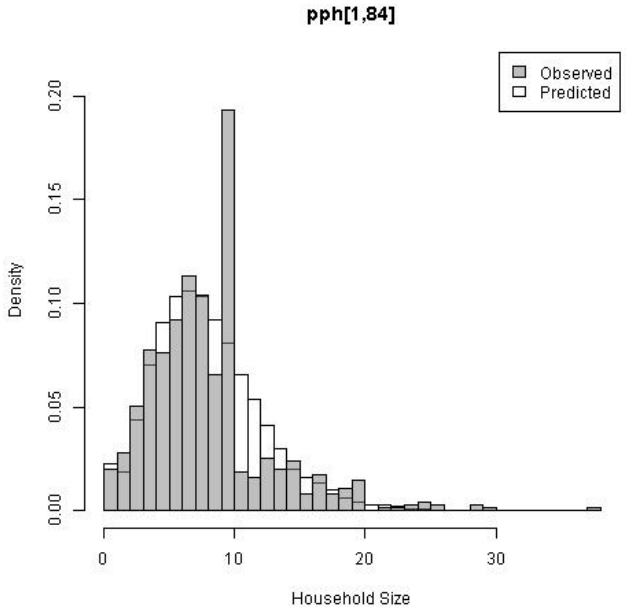

pph[1,8]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

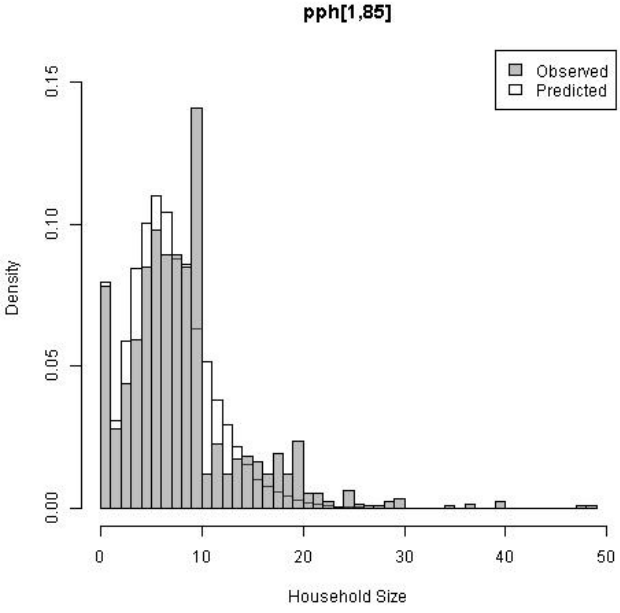

pph[1,80]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

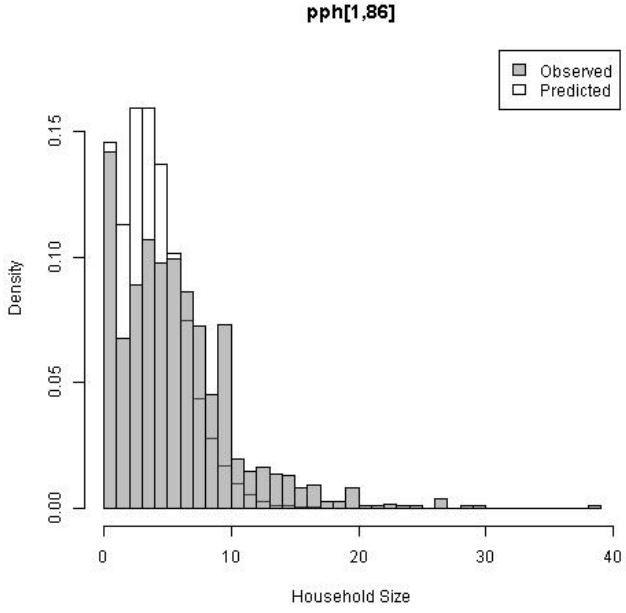

pph[1,81]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

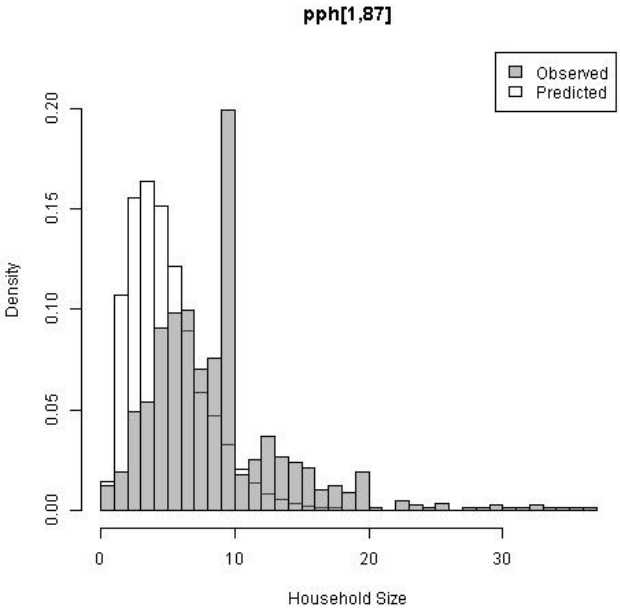

pph[1,82]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
50
Household Size

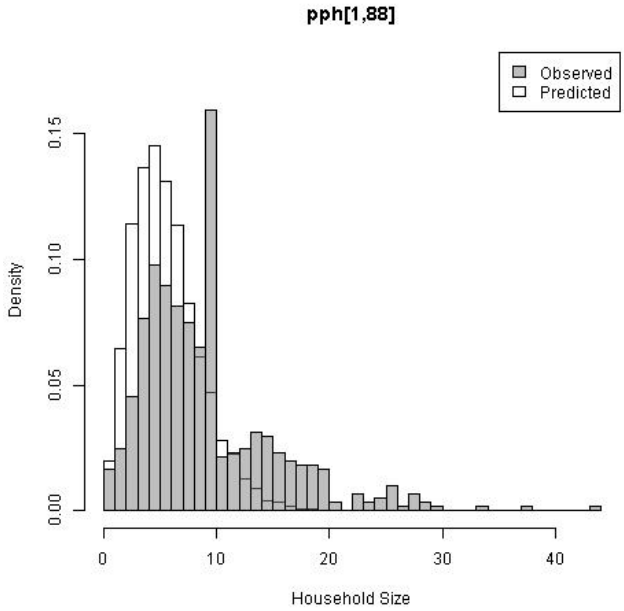

pph[1,83]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
Household Size

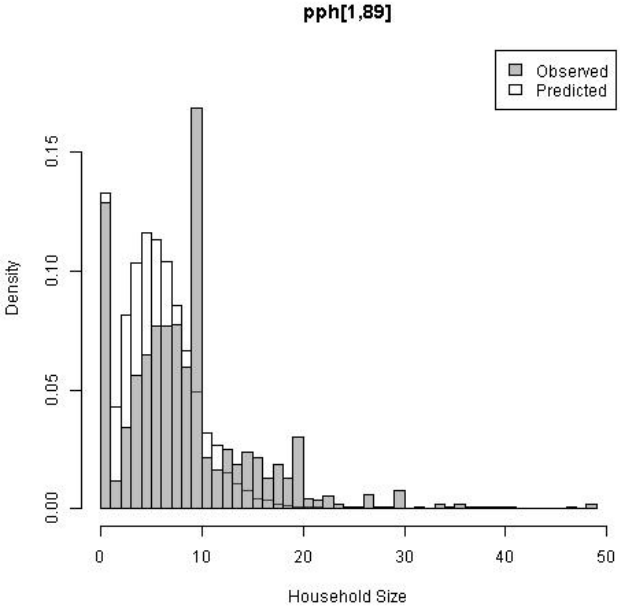

pph[1,84]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
Household Size

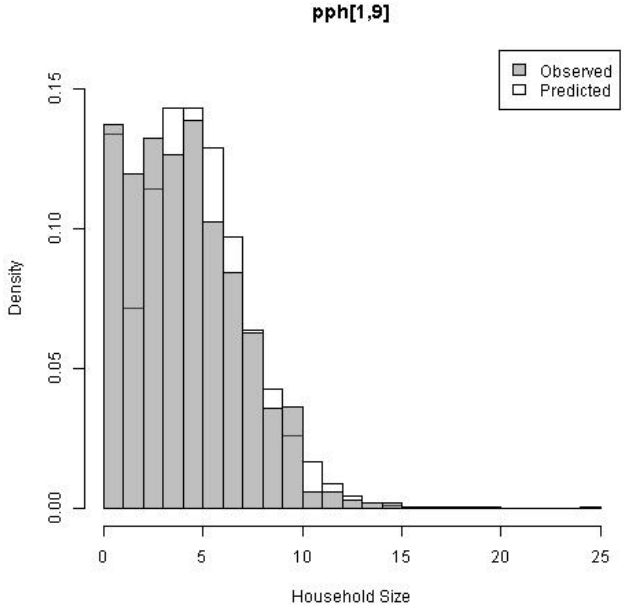

pph[1,85]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
50
Household Size

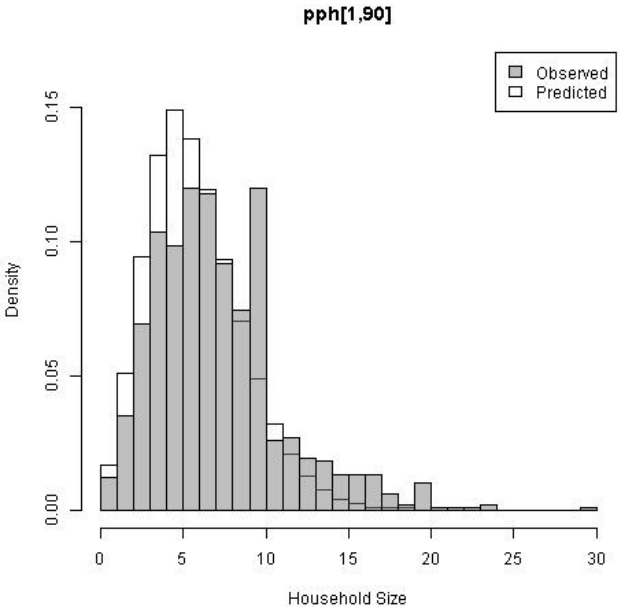

pph[1,86]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size

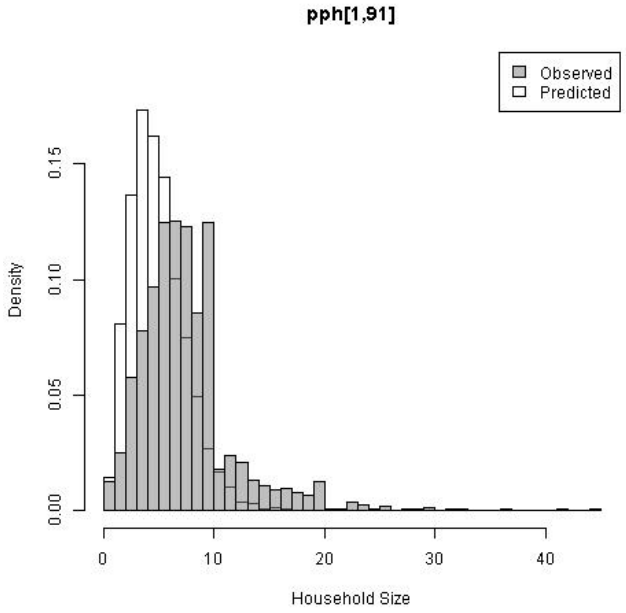

pph[1,87]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
Household Size

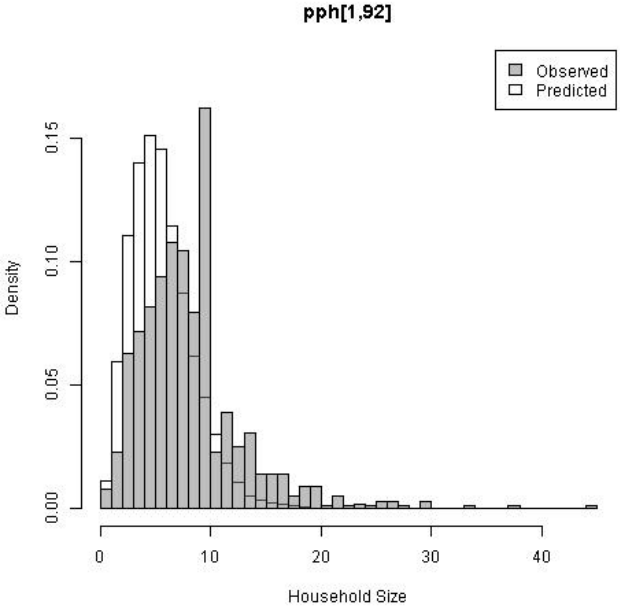

pph[1,88]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size

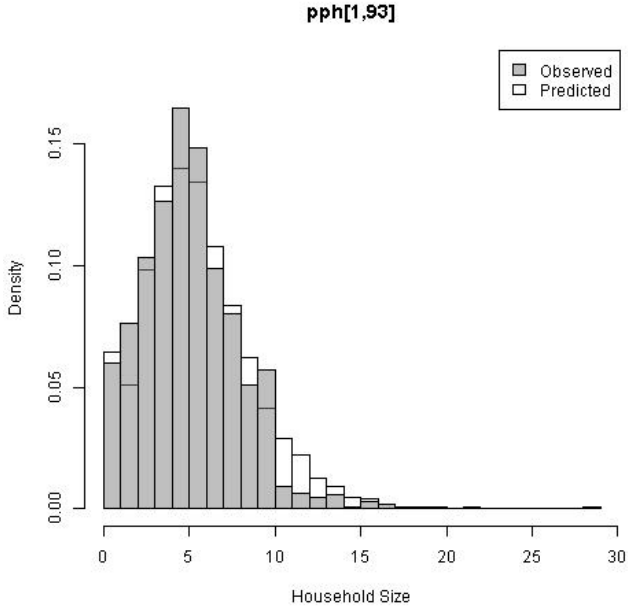

pph[1,89]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
50
Household Size

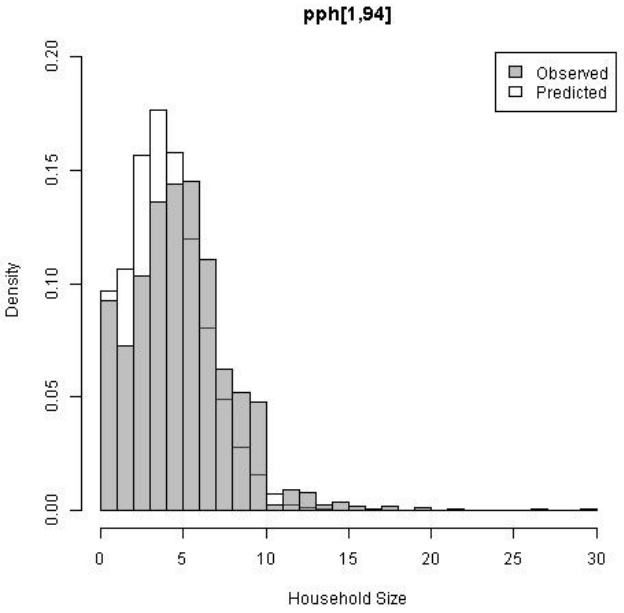

pph[1,9]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

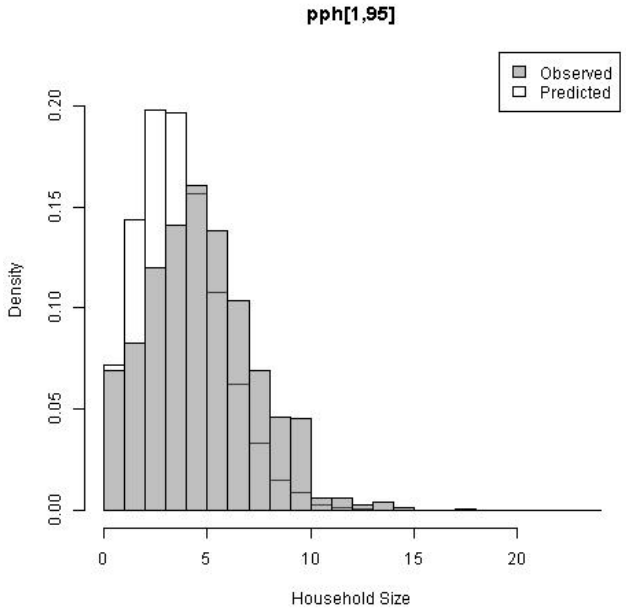

pph[1,90]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

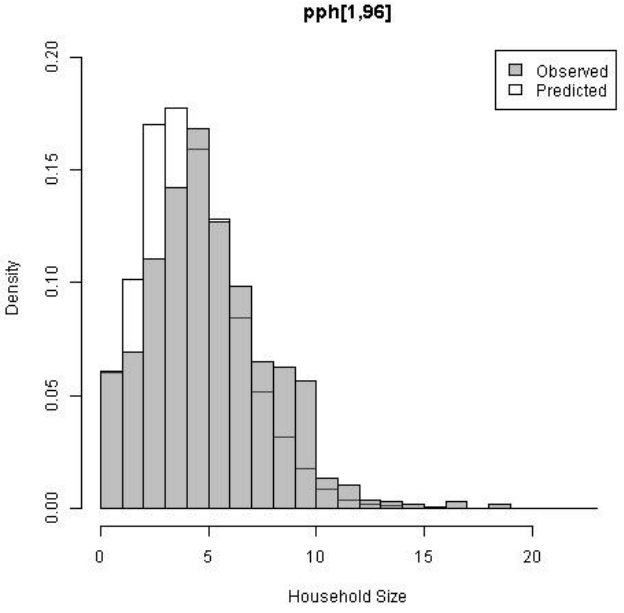

pph[1,91]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size

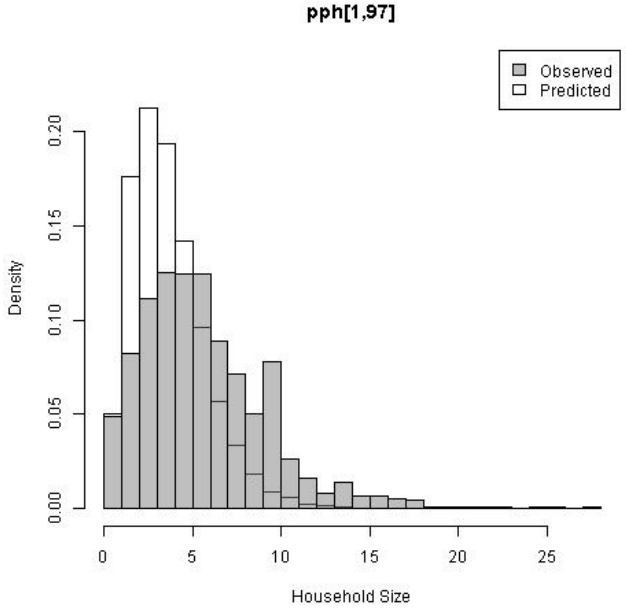

pph[1,92]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size

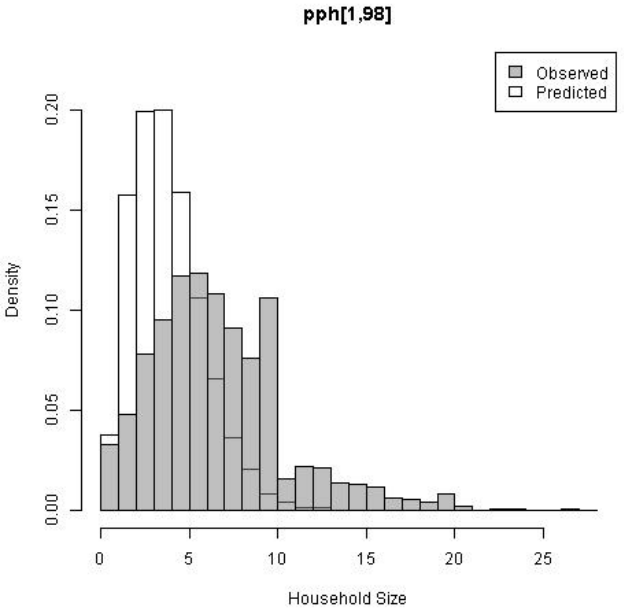

pph[1,93]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

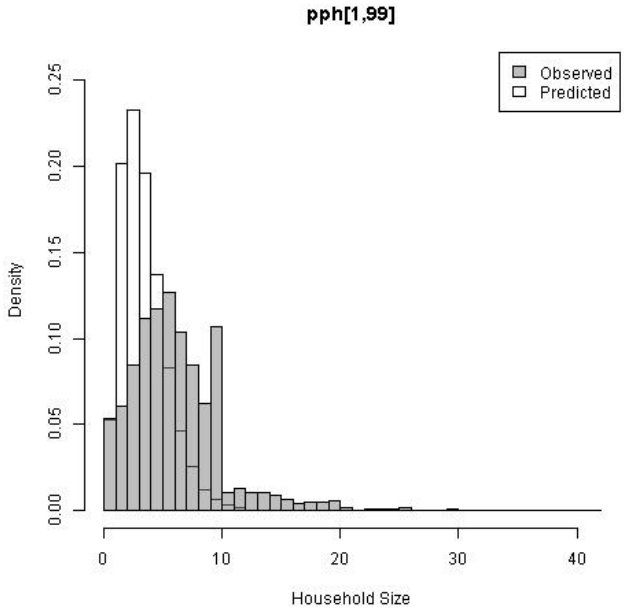

pph[1,94]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

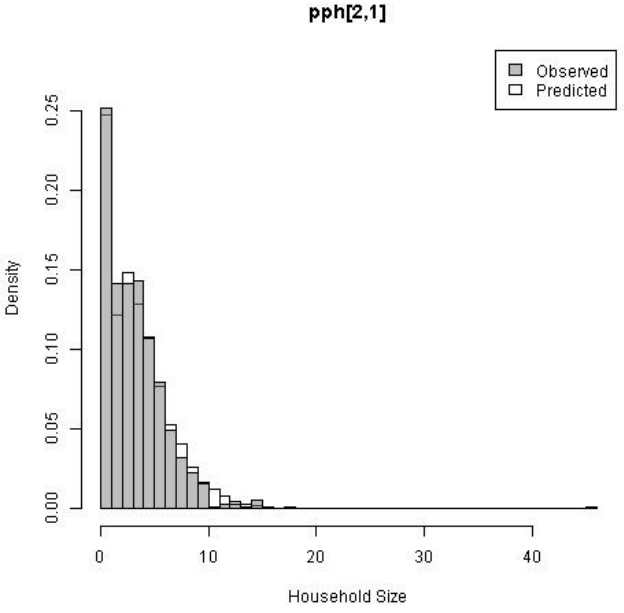

pph[1,95]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

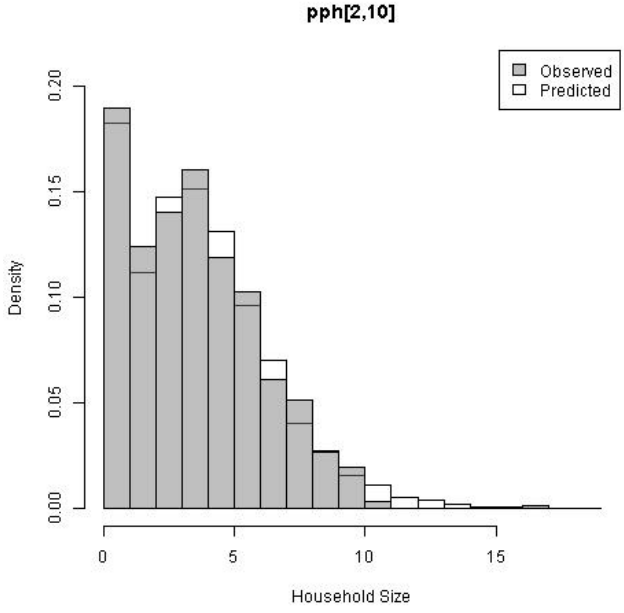

pph[1,96]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

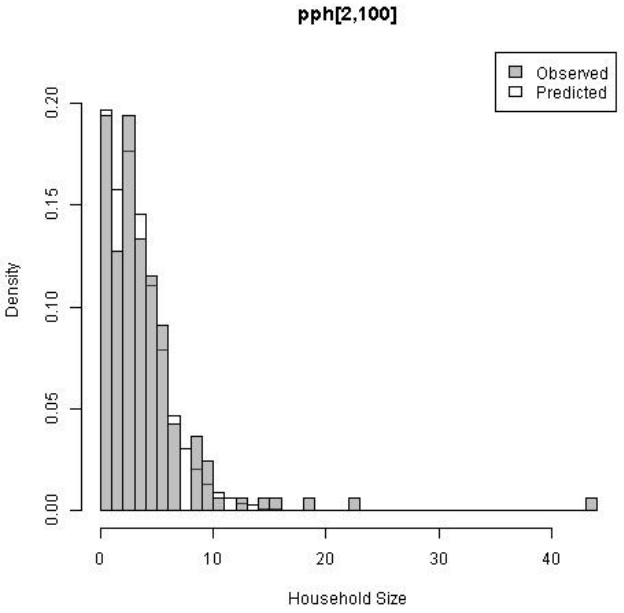

pph[1,97]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

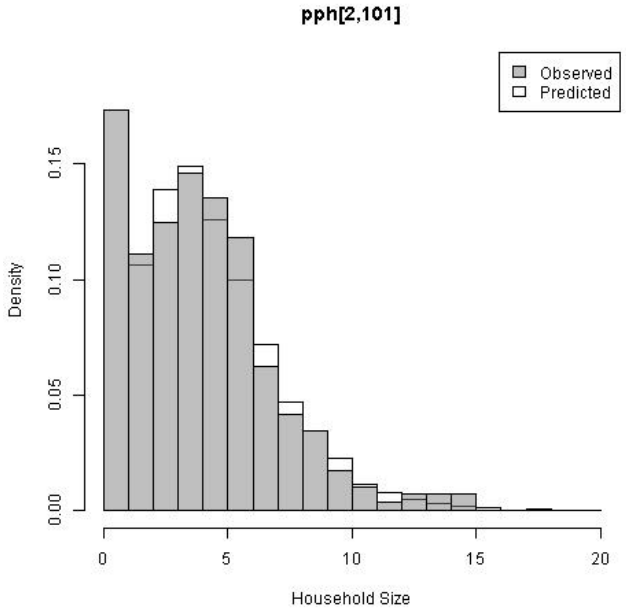

pph[1,98]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

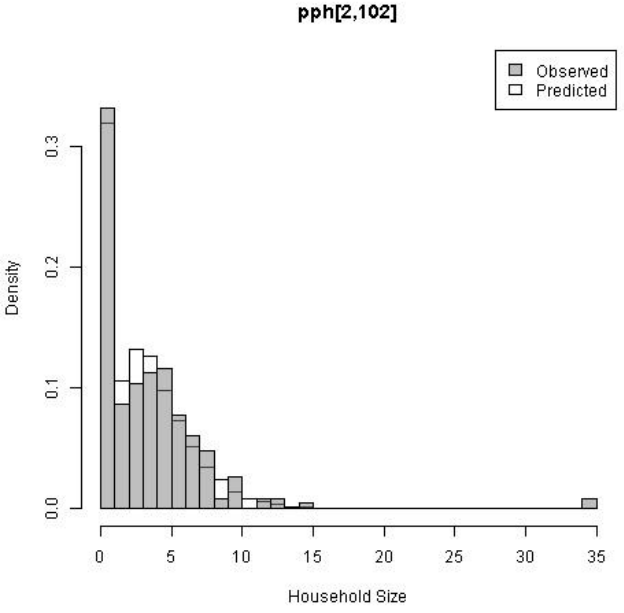

pph[1,99]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Household Size

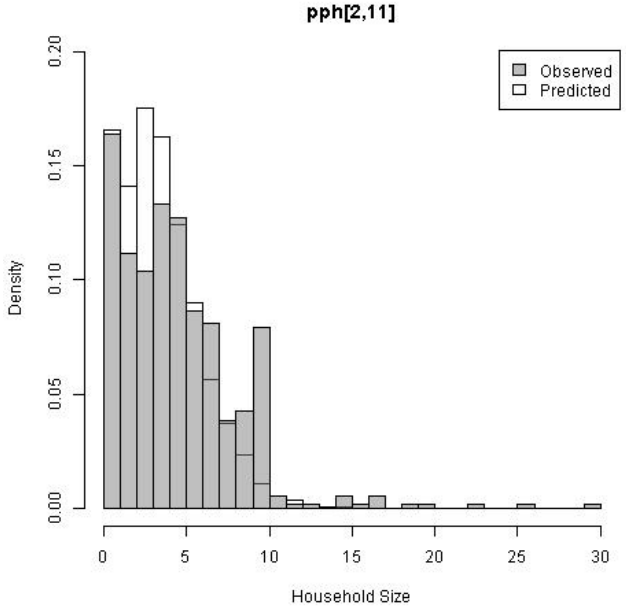

pph[2,1]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Household Size

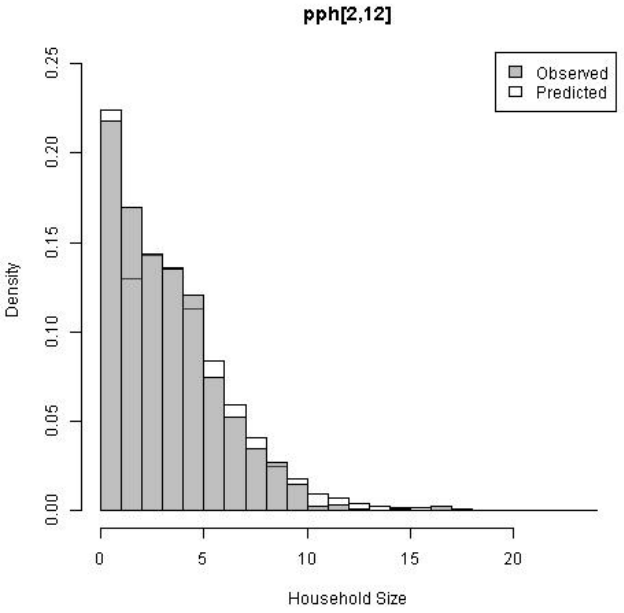

pph[2,10]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

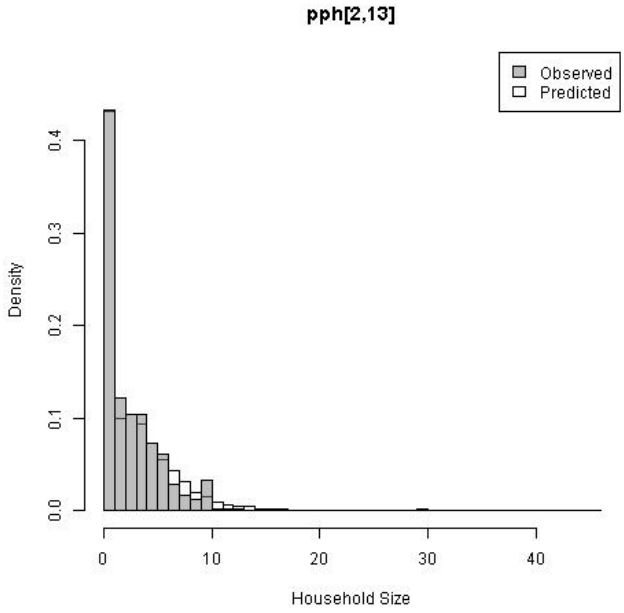

pph[2,100]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
Household Size

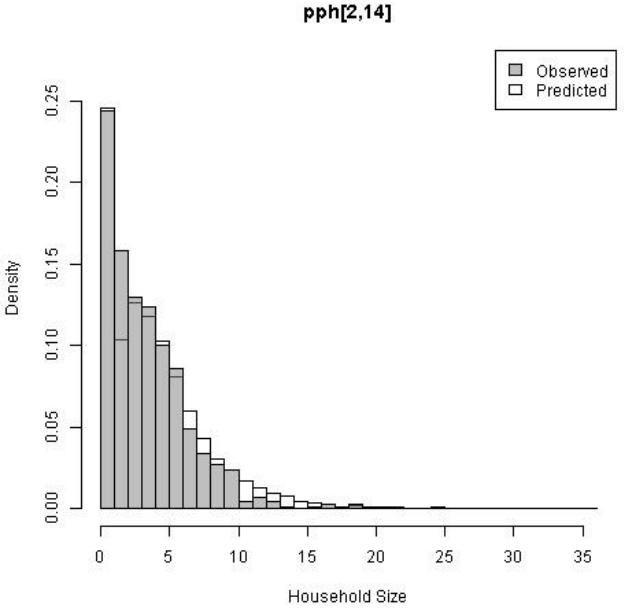

pph[2,101]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

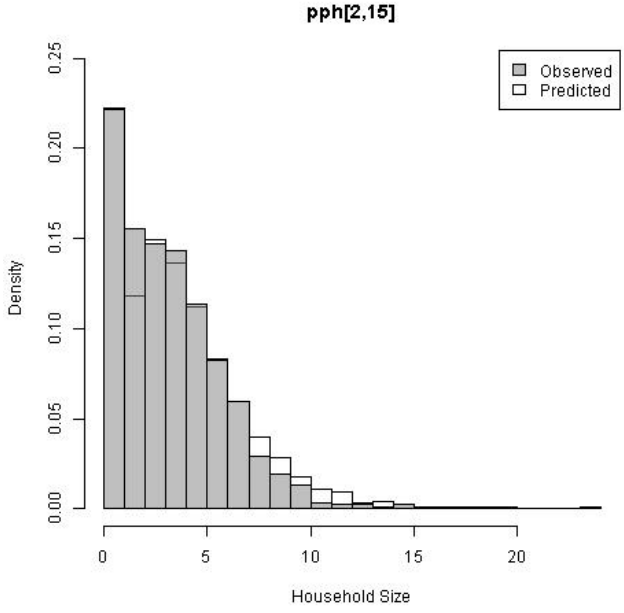

pph[2,102]
Observed
Predicted
Density
0.0
0.1
0.2
0.3
0
5
10
15
20
25
30
35
Household Size

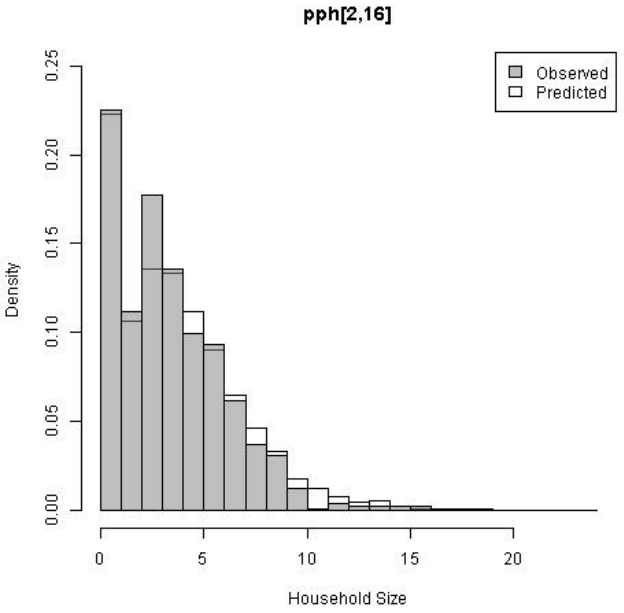

pph[2,11]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

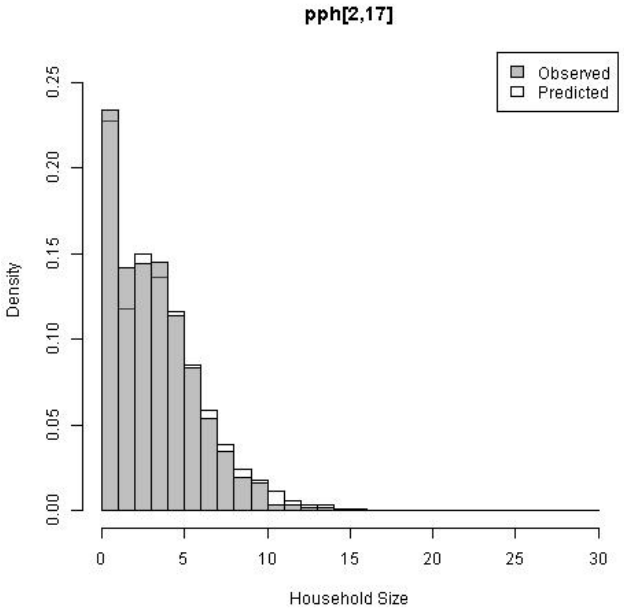

pph[2,12]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

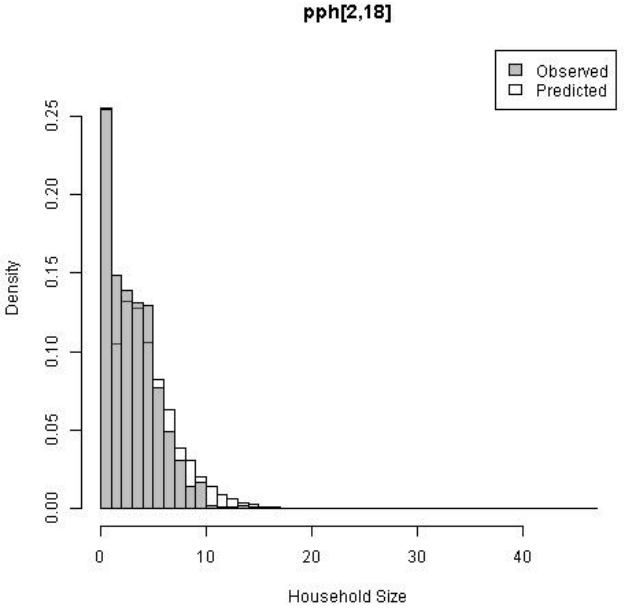

pph[2,13]
Observed
Predicted
Density
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
Household Size

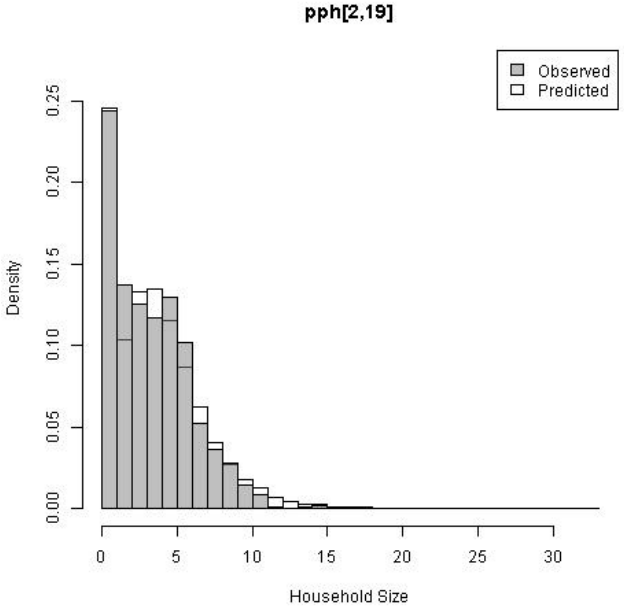

pph[2,14]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
35
Household Size

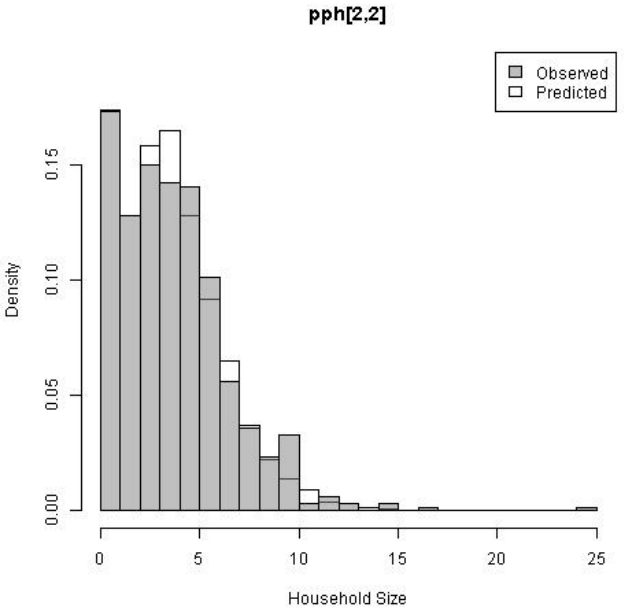

pph[2,15]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

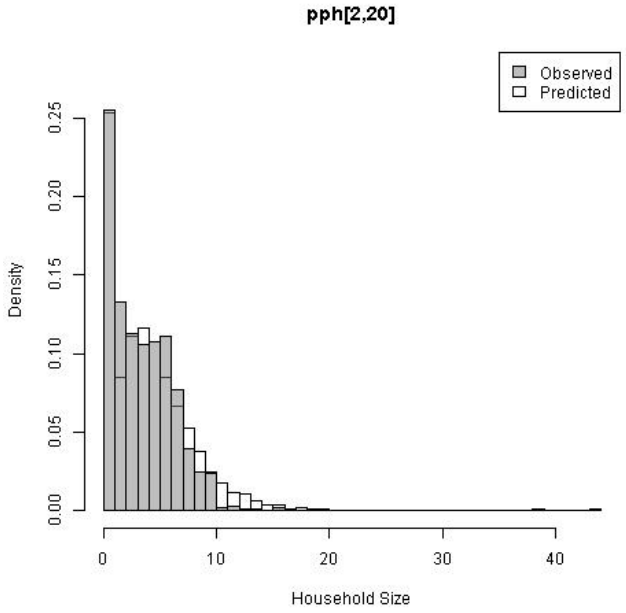

pph[2,16]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

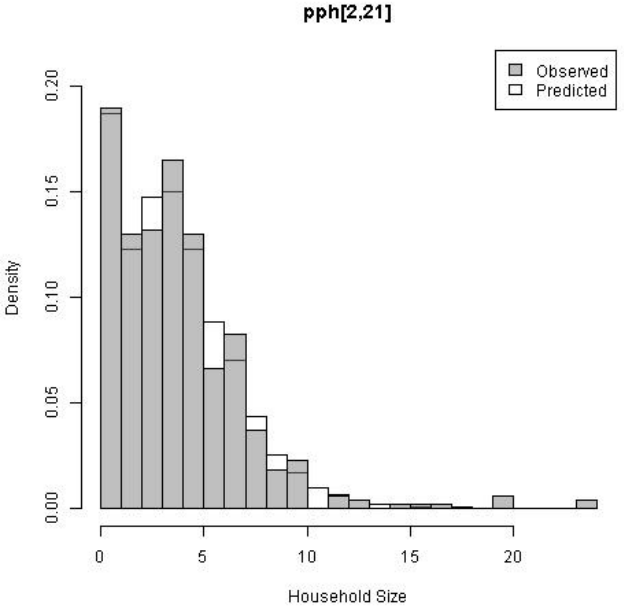

pph[2,17]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
Household Size

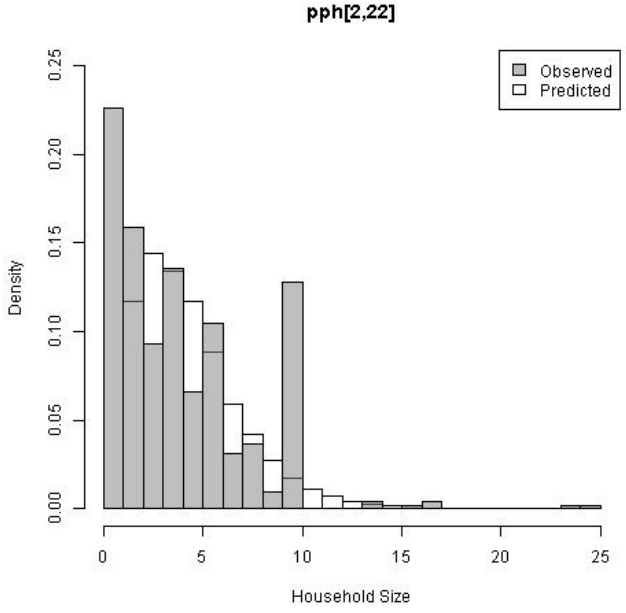

pph[2,18]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Household Size

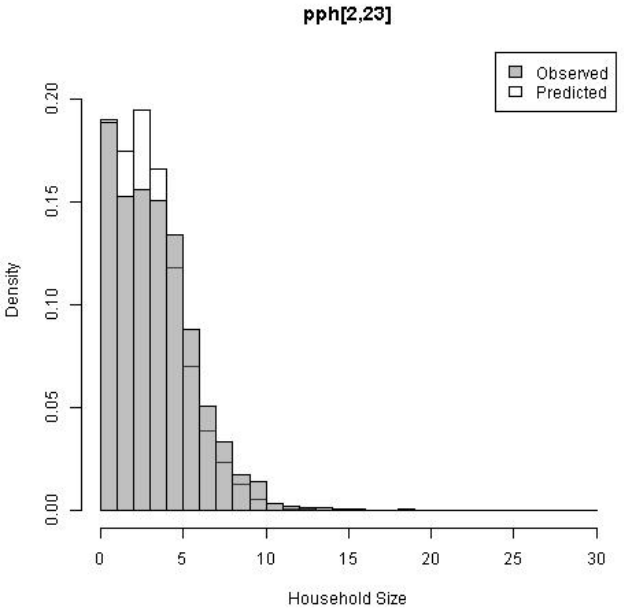

pph[2,19]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
Household Size

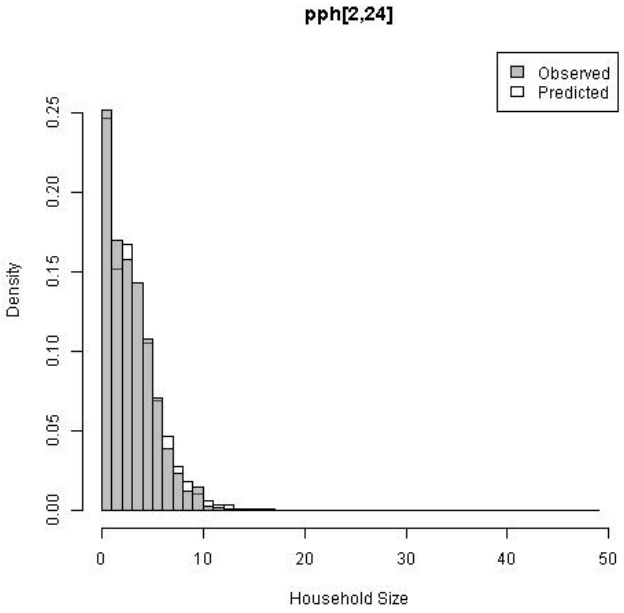

pph[2,2]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

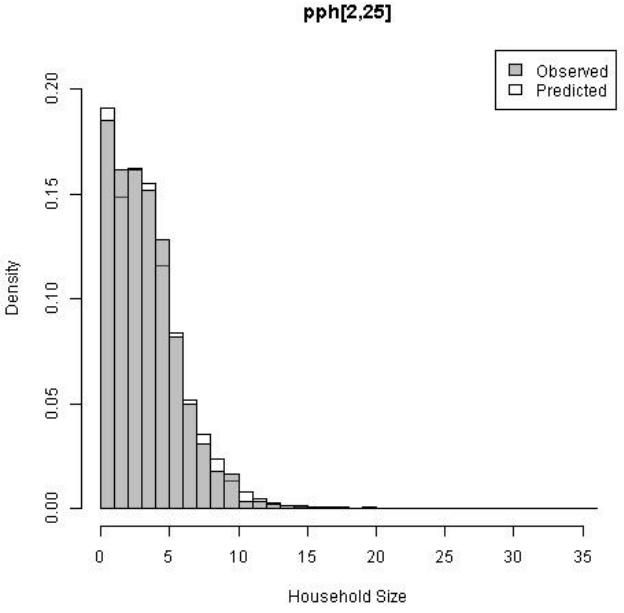

pph[2,20]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Household Size

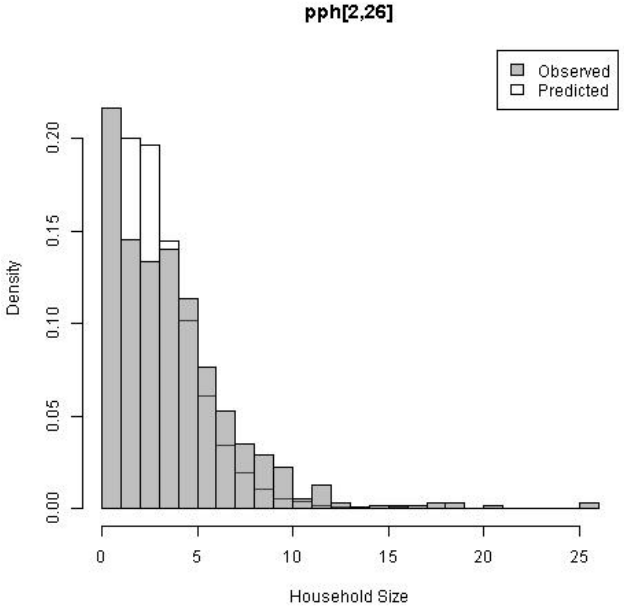

pph[2,21]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

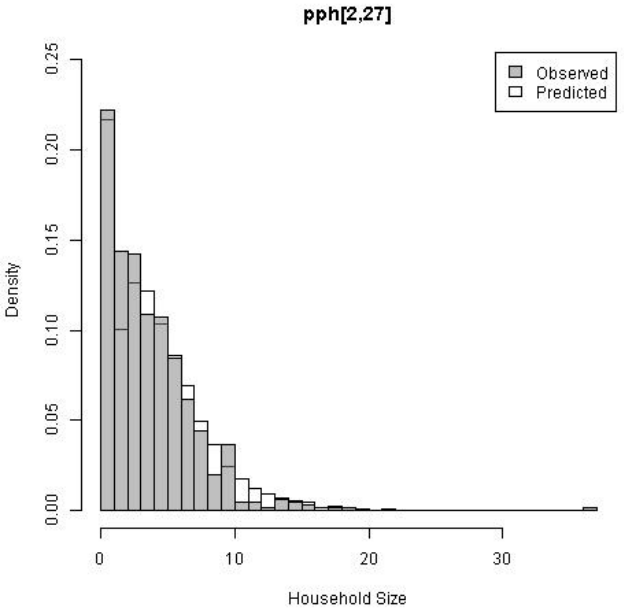

pph[2,22]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
Household Size

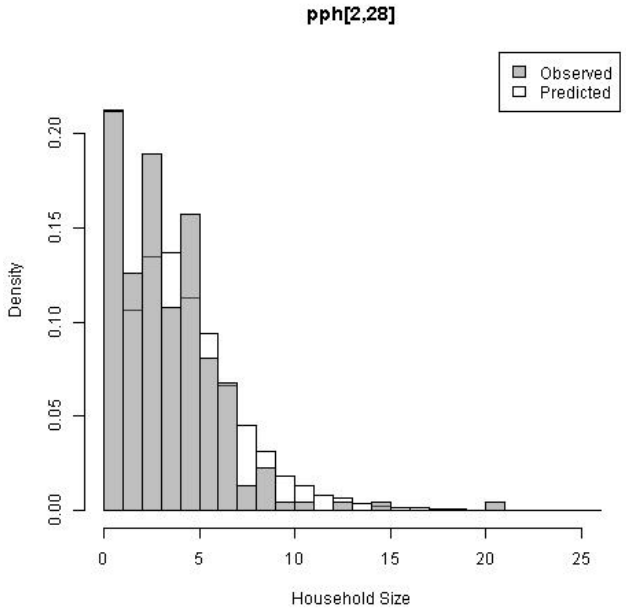

pph[2,23]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

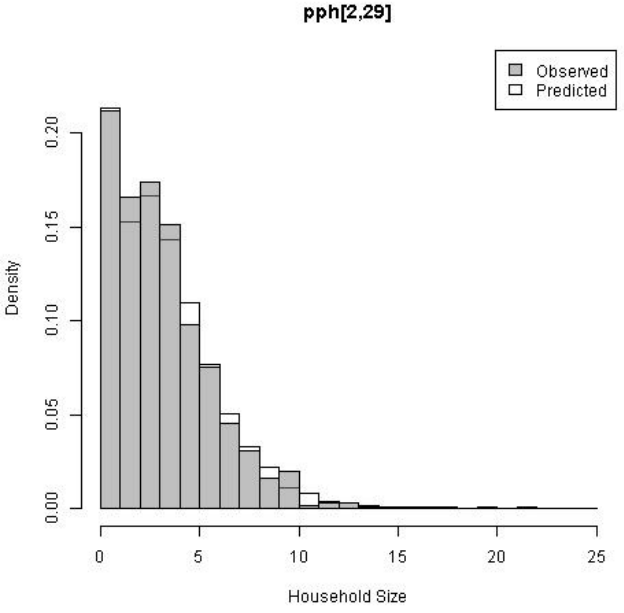

pph[2,24]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
50
Household Size

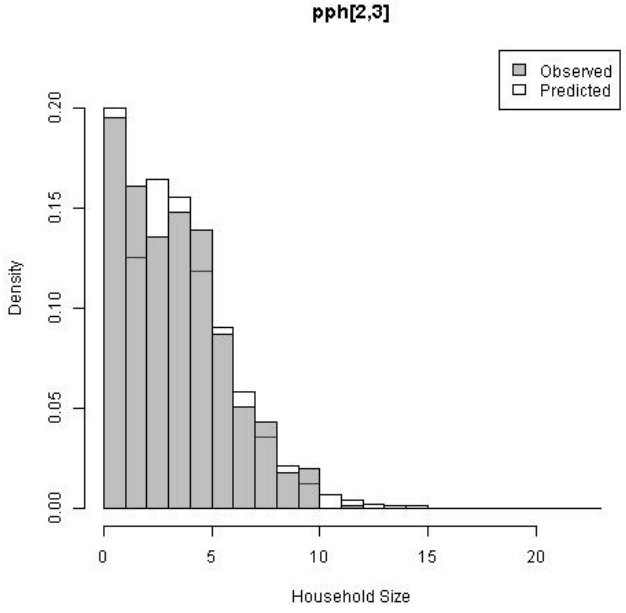

pph[2,25]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
35
Household Size

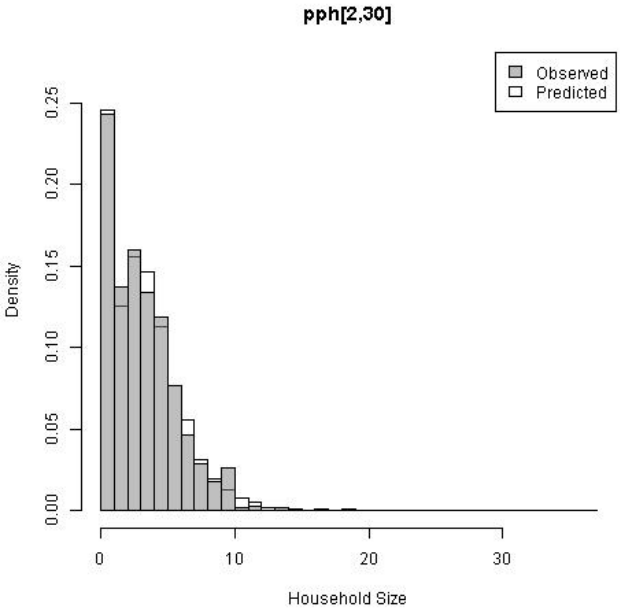

pph[2,26]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

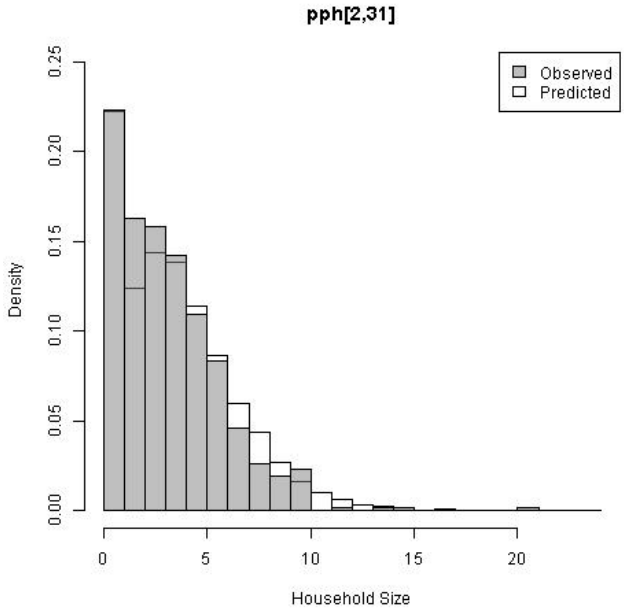

pph[2,27]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
Household Size

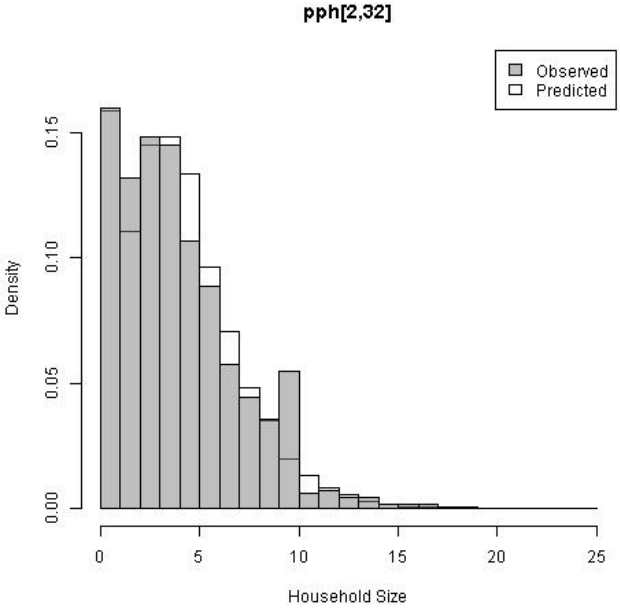

pph[2,28]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

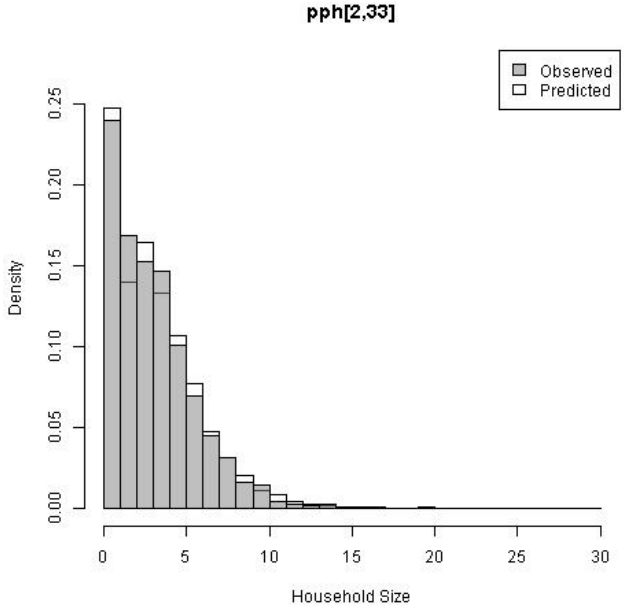

pph[2,29]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

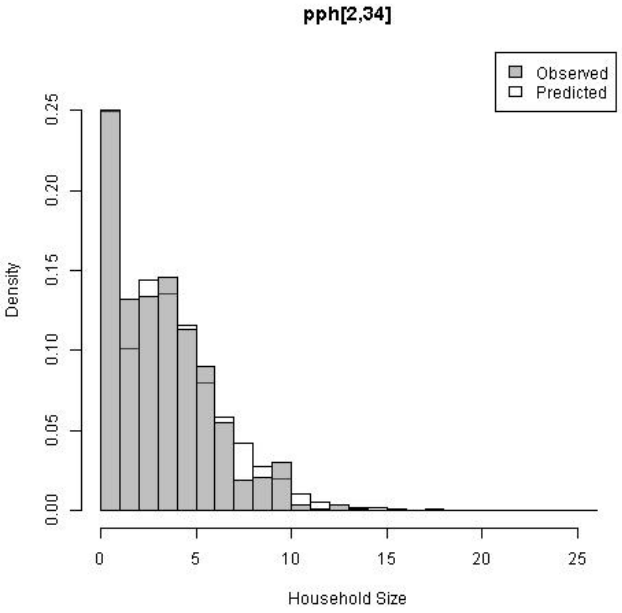

pph[2,3]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

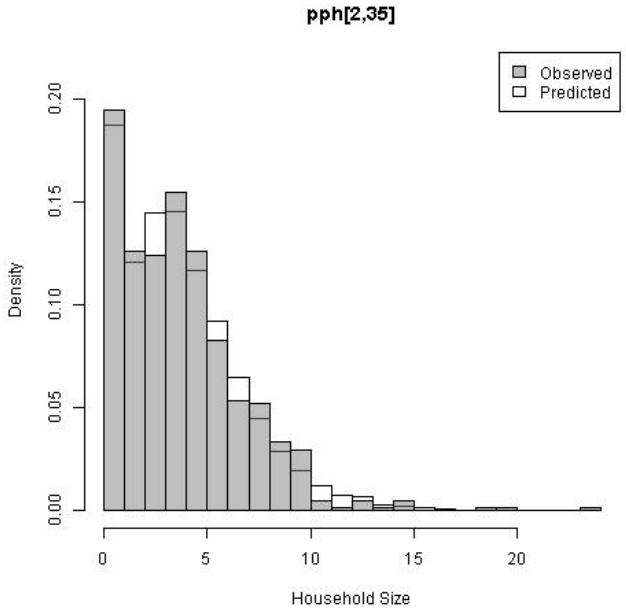

pph[2,30]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
Household Size

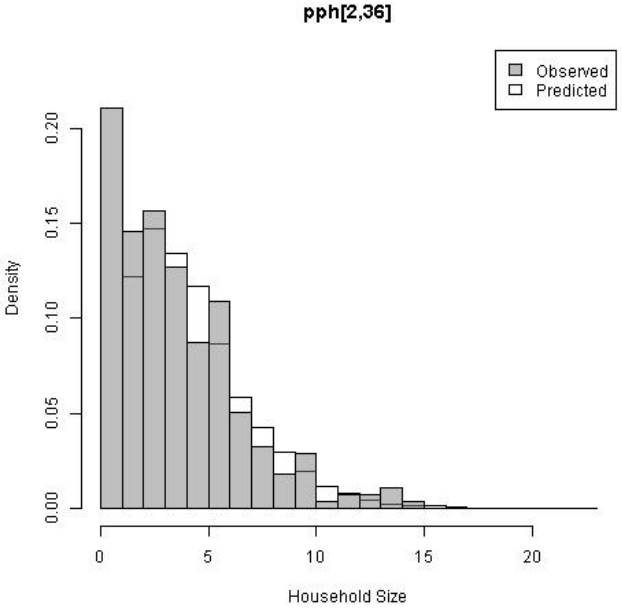

pph[2,31]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

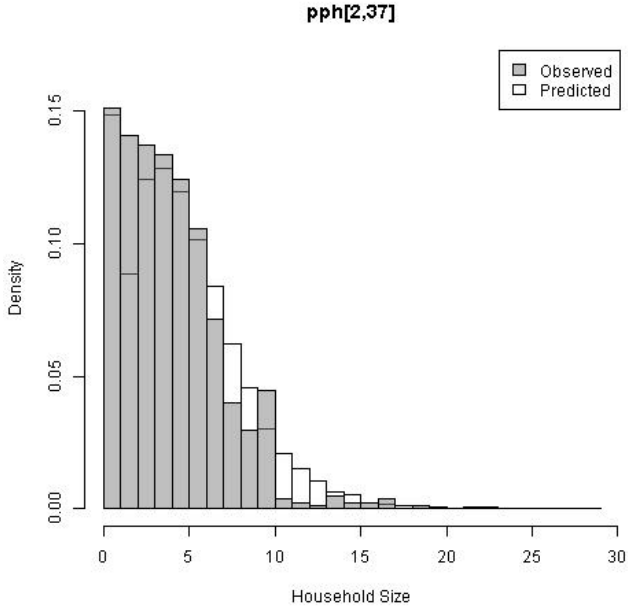

pph[2,32]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

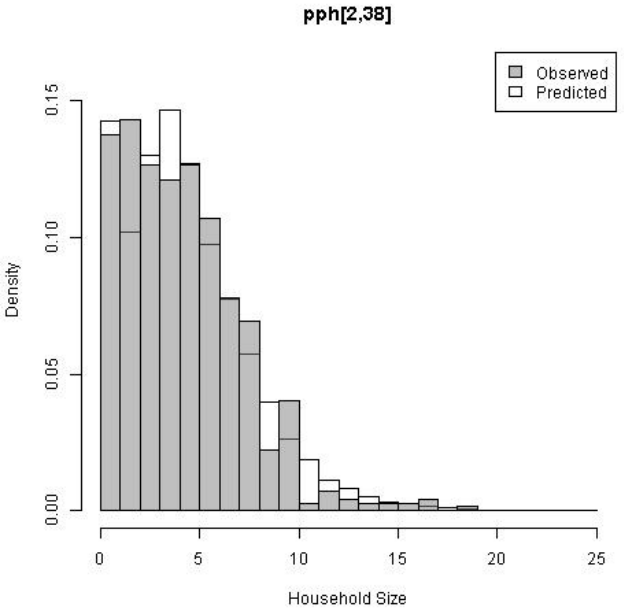

pph[2,33]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
Household Size

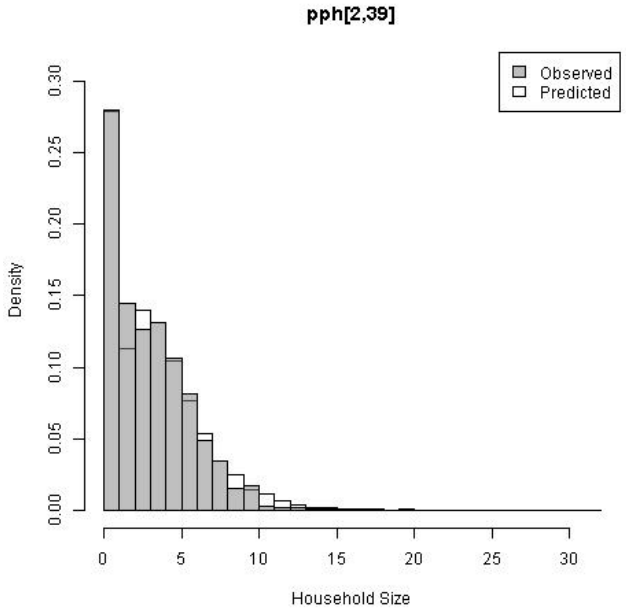

pph[2,34]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
Household Size

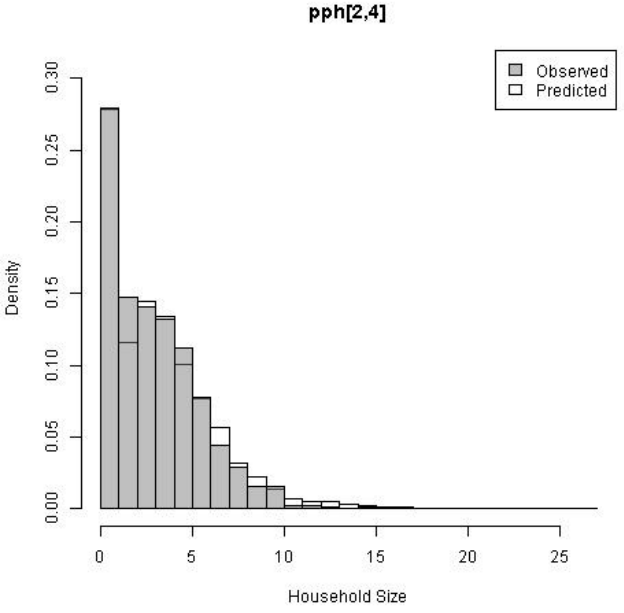

pph[2,35]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

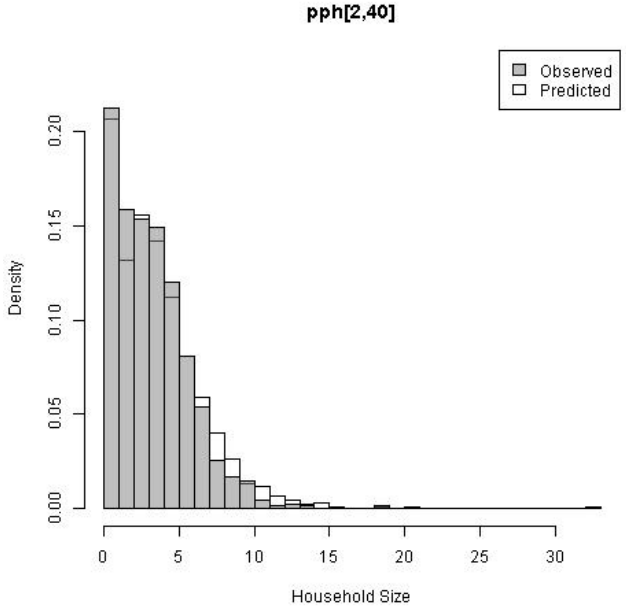

pph[2,36]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

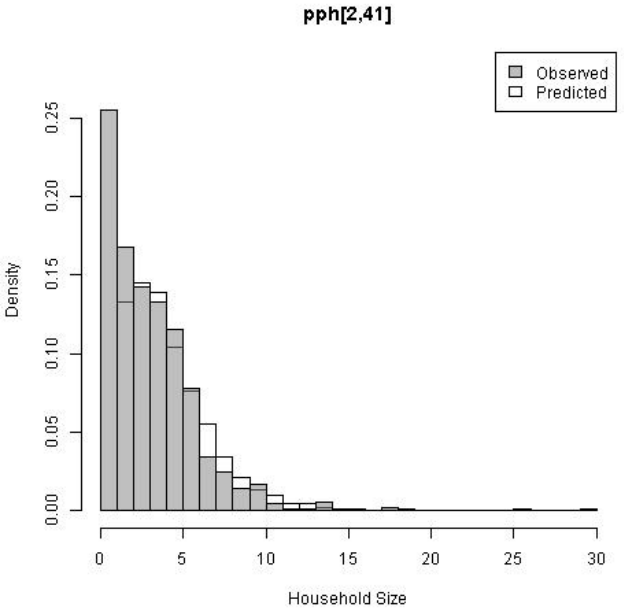

pph[2,37]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

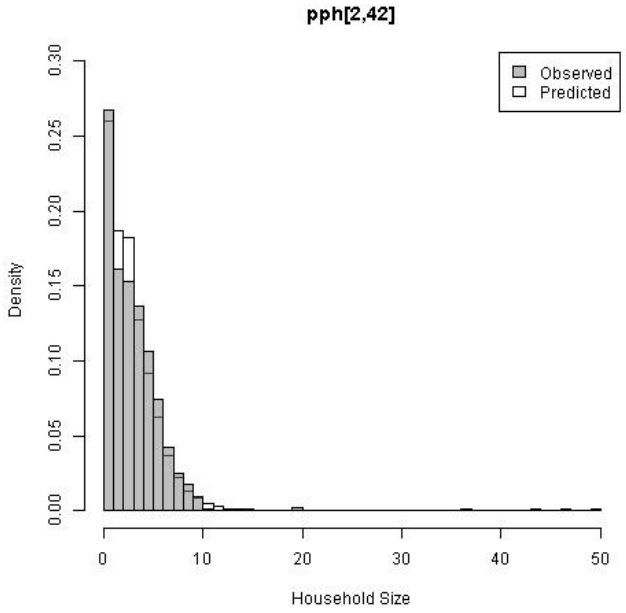

pph[2,38]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

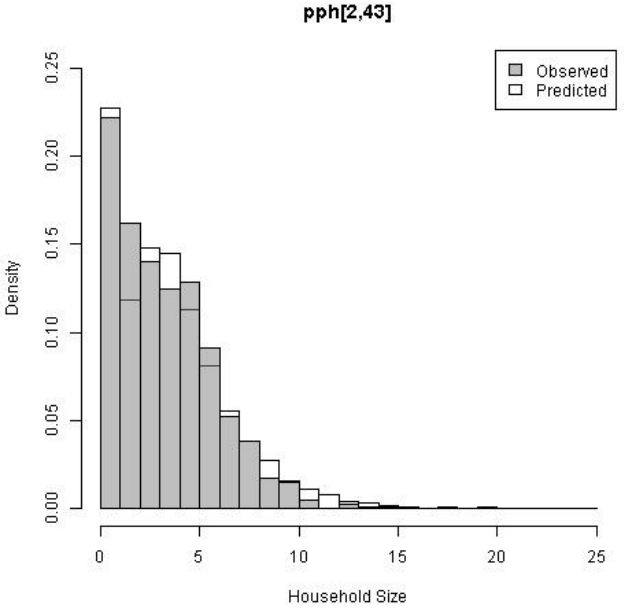

pph[2,39]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20
25
30
Household Size

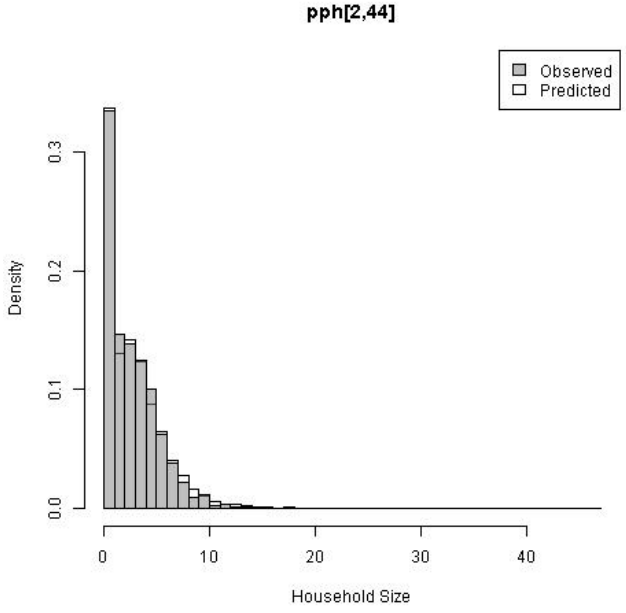

pph[2,4]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20
25
Household Size

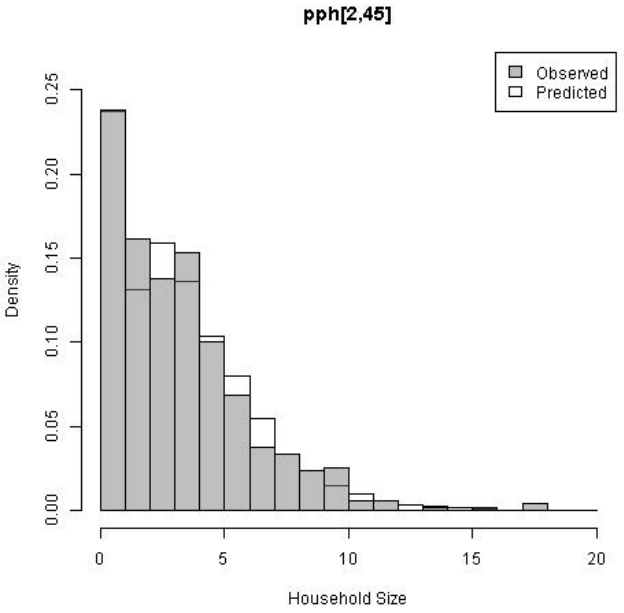

pph[2,40]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

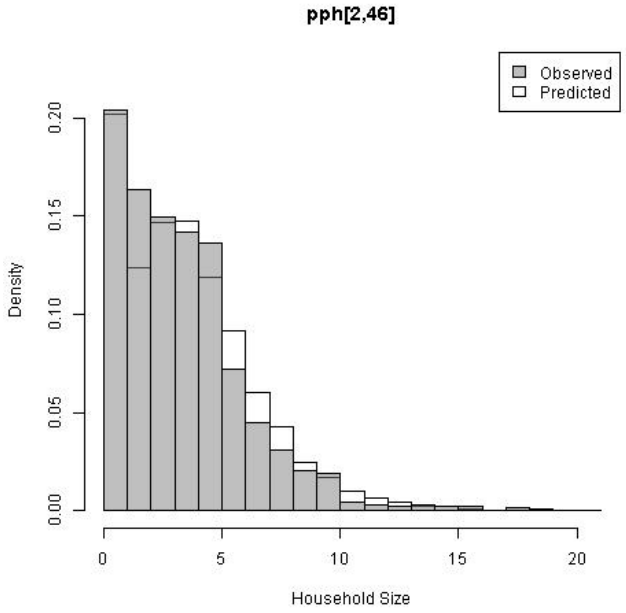

pph[2,41]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
Household Size

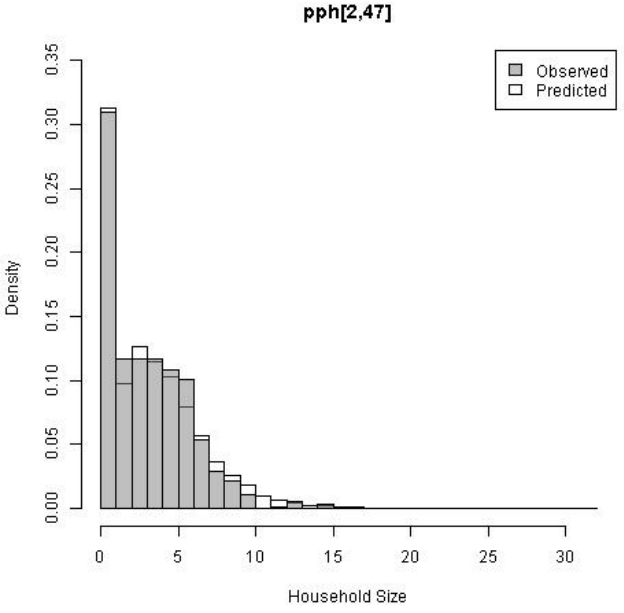

pph[2,42]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
Household Size

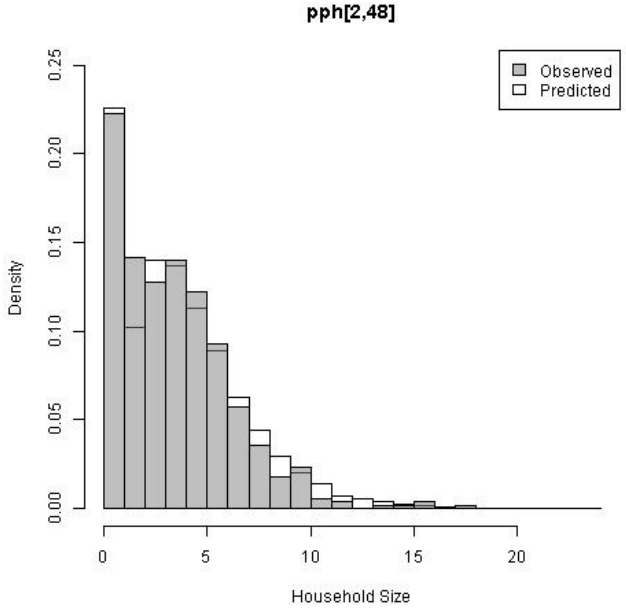

pph[2,43]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
Household Size

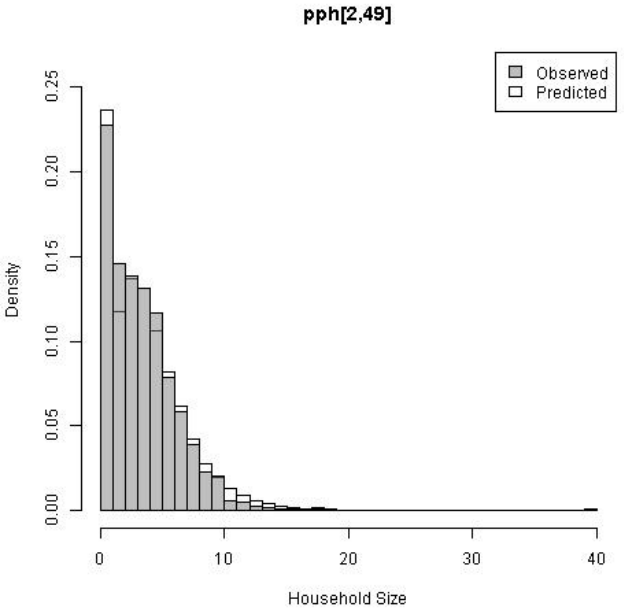

pph[2,44]
Observed
Predicted
Density
0.0
0.1
0.2
0.3
0
10
20
30
40
Household Size

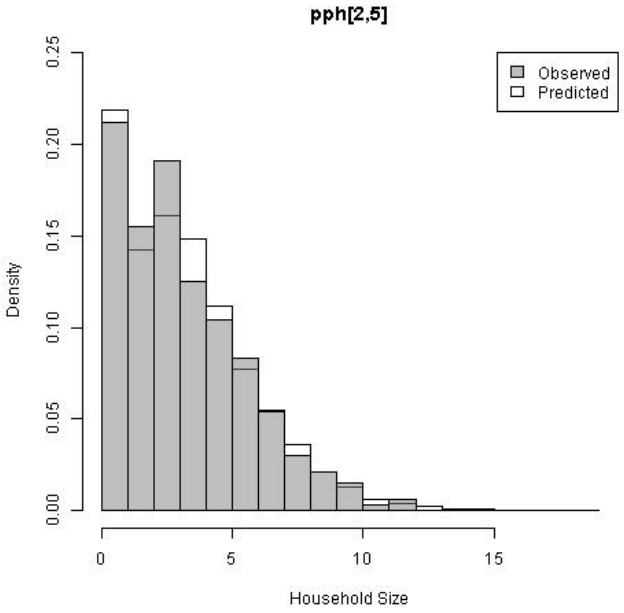

pph[2,45]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

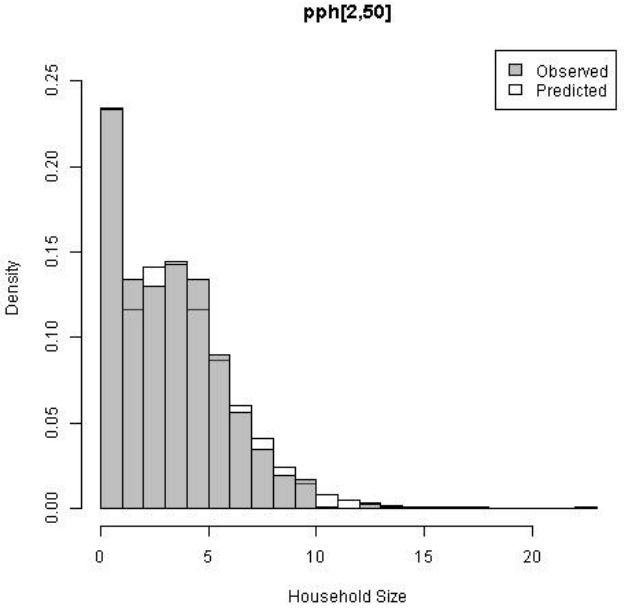

pph[2,46]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

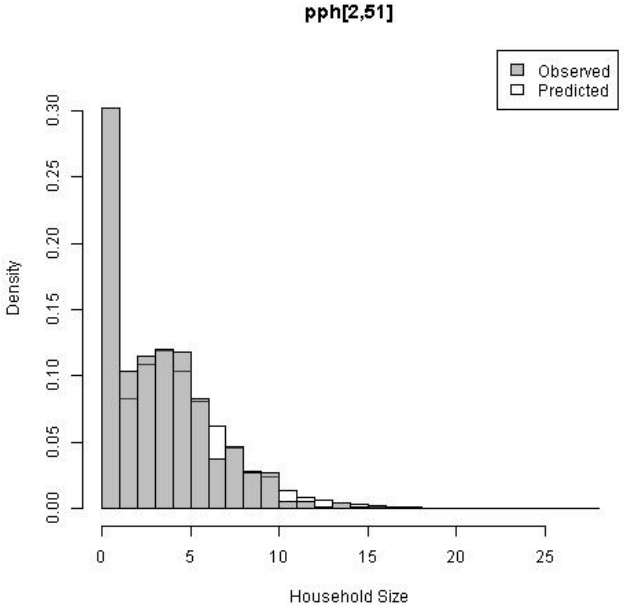

pph[2,47]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
5
10
15
20
25
30
Household Size

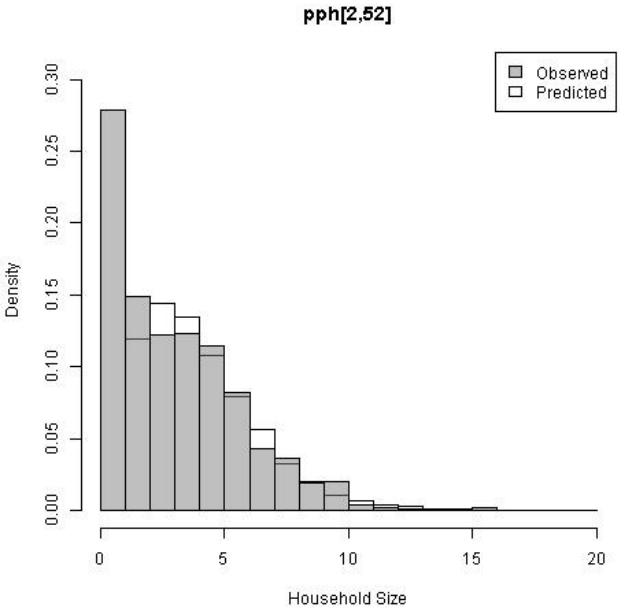

pph[2,48]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

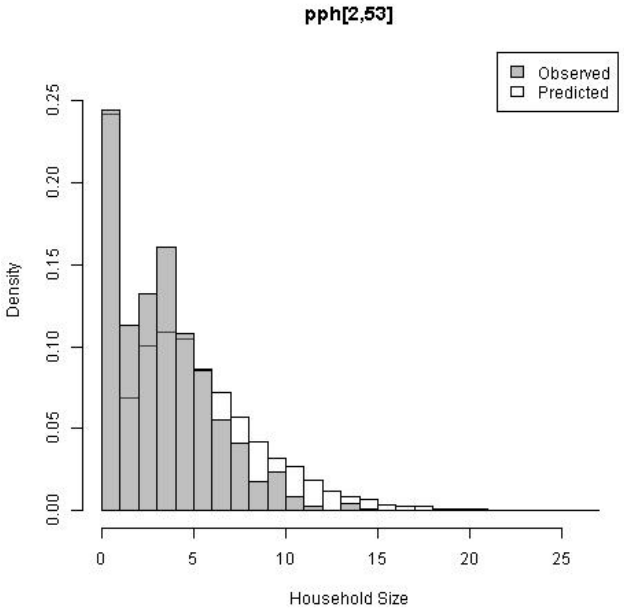

pph[2,49]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Household Size

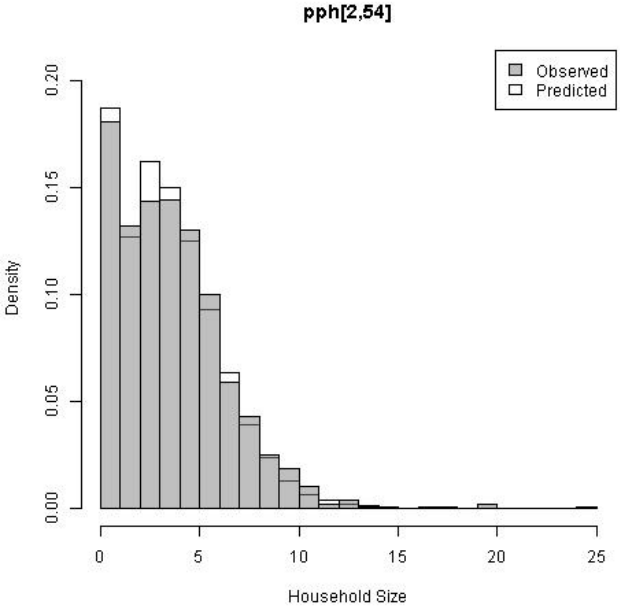

pph[2,5]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
Household Size

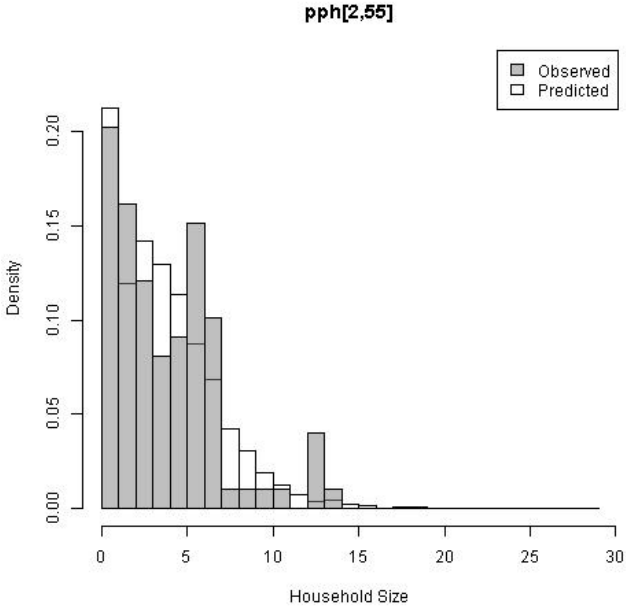

pph[2,50]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

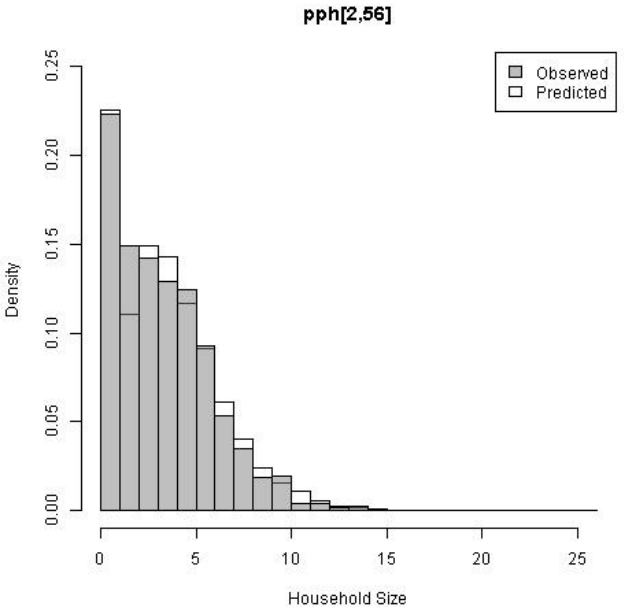

pph[2,51]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20
25
Household Size

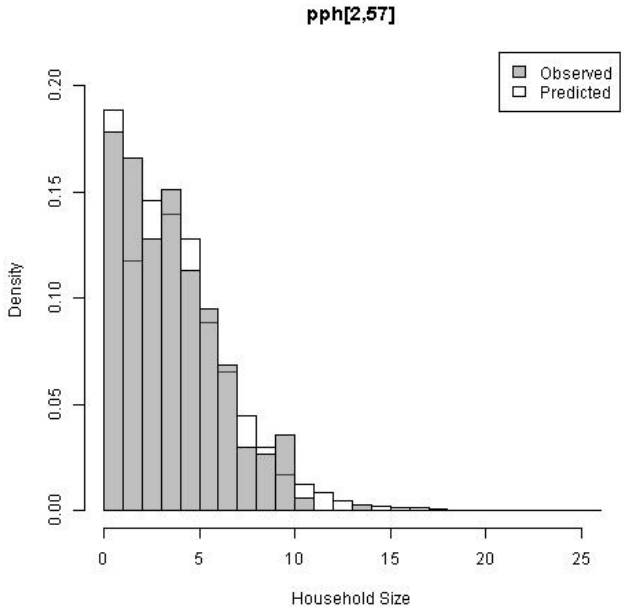

pph[2,52]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20
Household Size

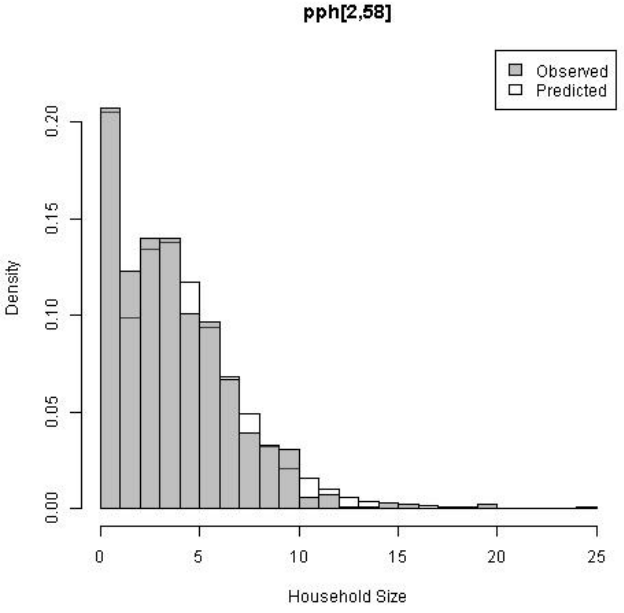

pph[2,53]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
Household Size

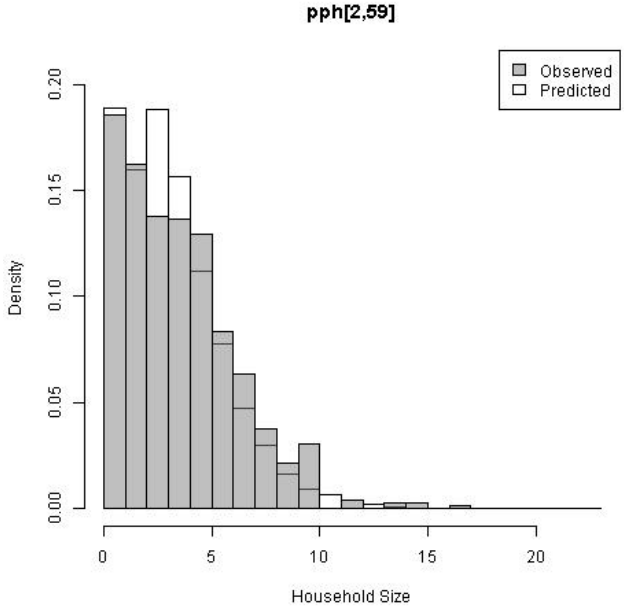

pph[2,54]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

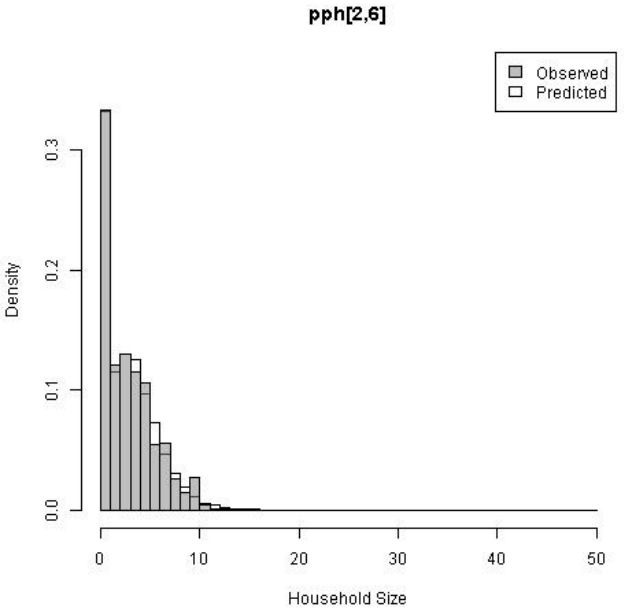

pph[2,55]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

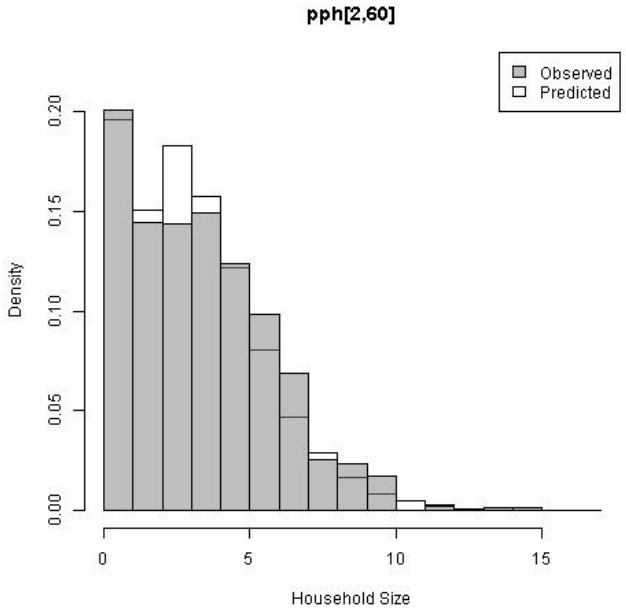

pph[2,56]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
Household Size

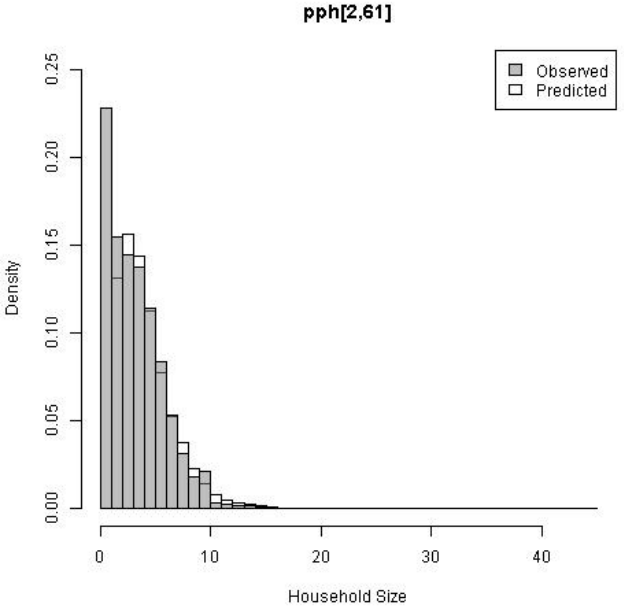

pph[2,57]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

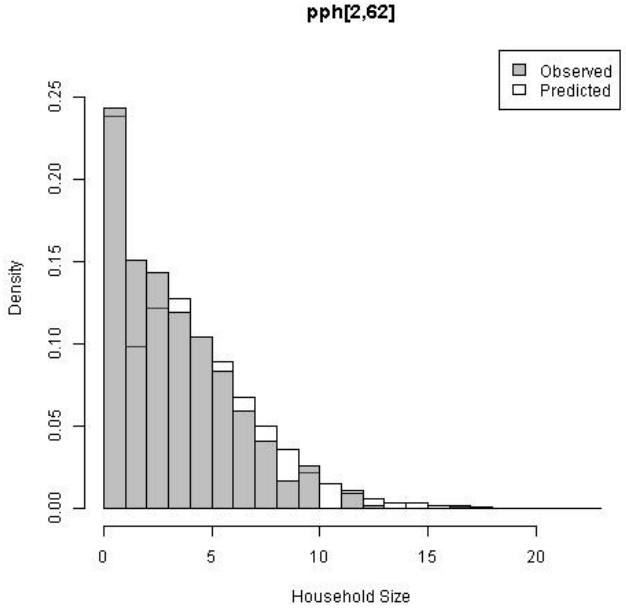

pph[2,58]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

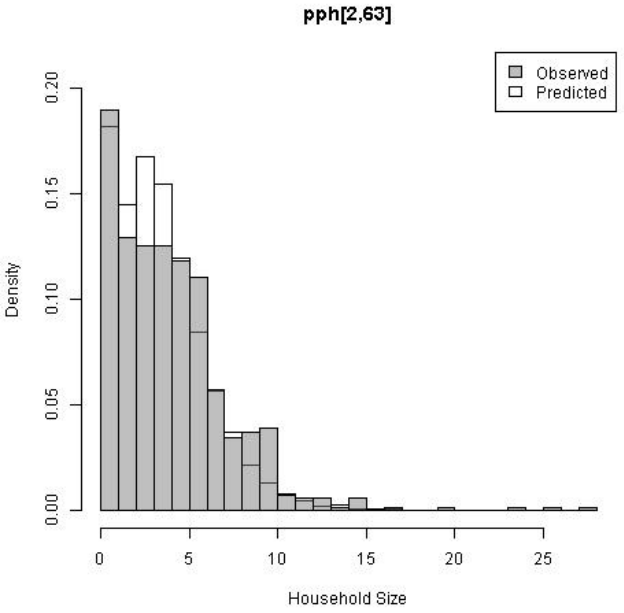

pph[2,59]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

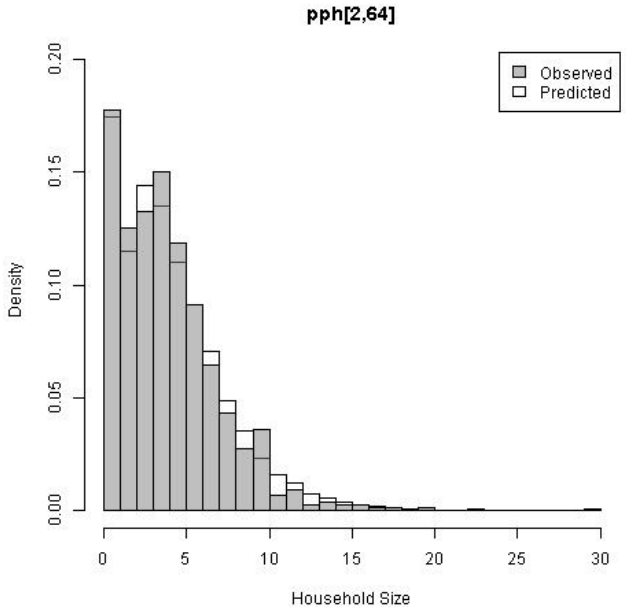

pph[2,6]
Observed
Predicted
Density
0.0
0.1
0.2
0.3
0
10
20
30
40
50
Household Size

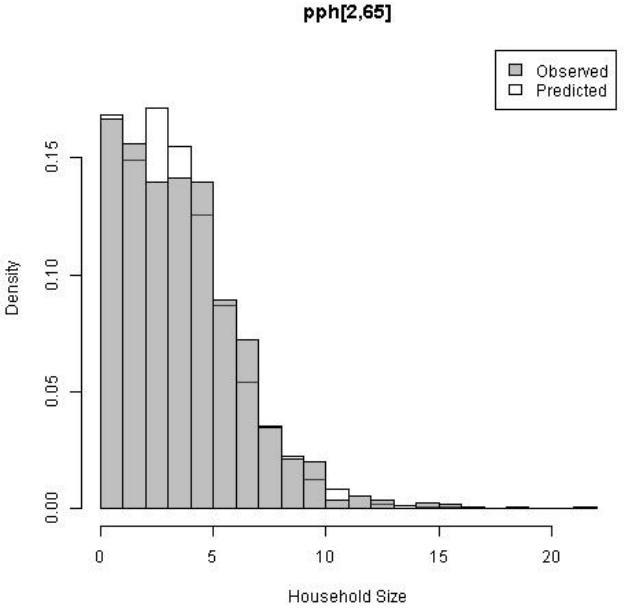

pph[2,60]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size

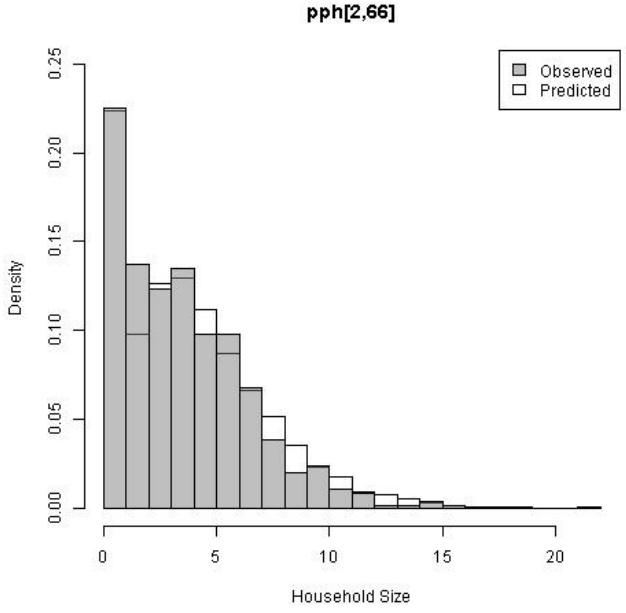

pph[2,61]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Household Size

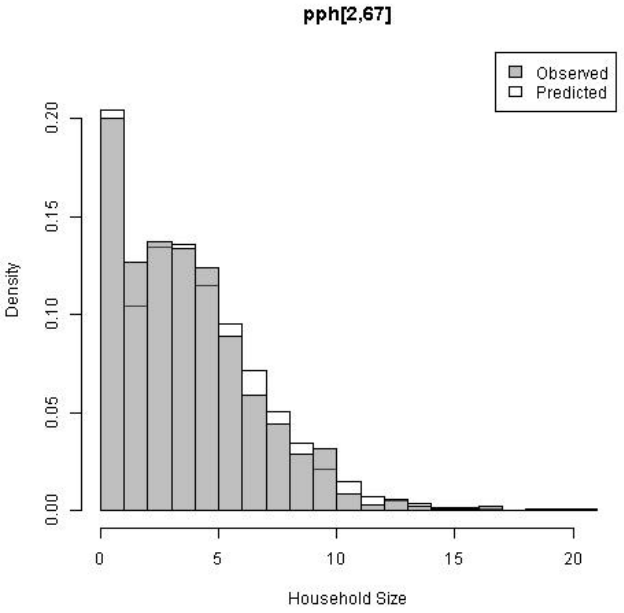

pph[2,62]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

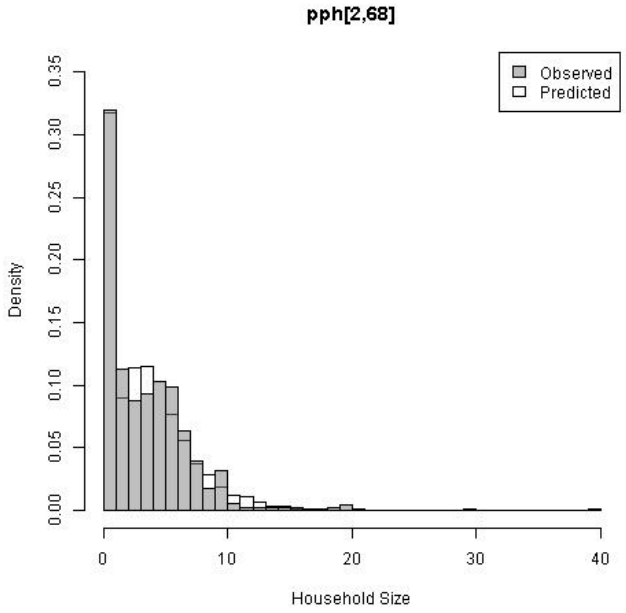

pph[2,63]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

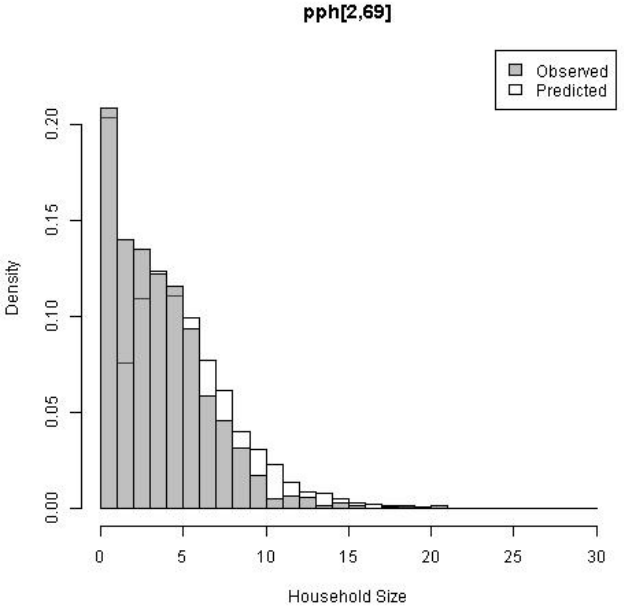

pph[2,64]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

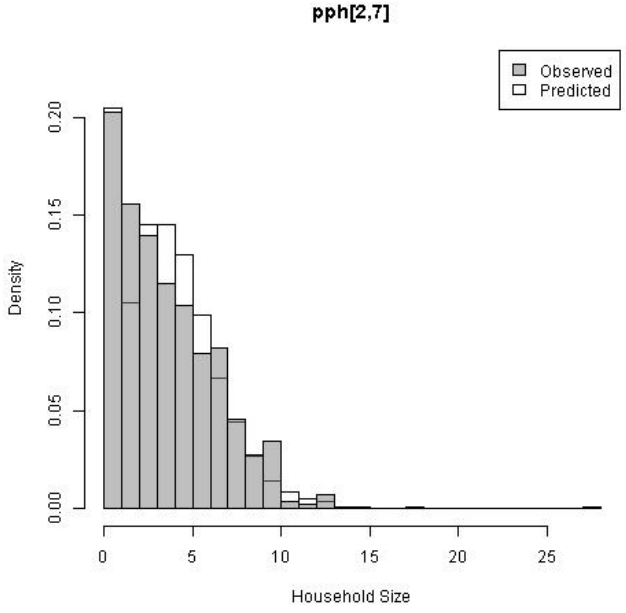

pph[2,65]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Household Size

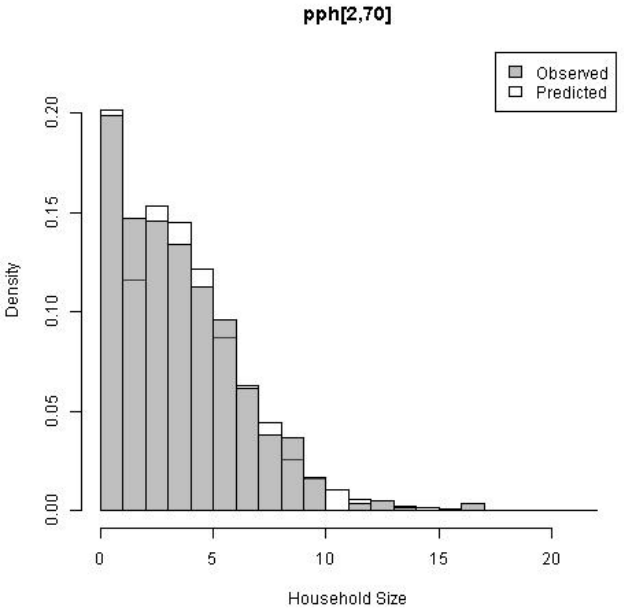

pph[2,66]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

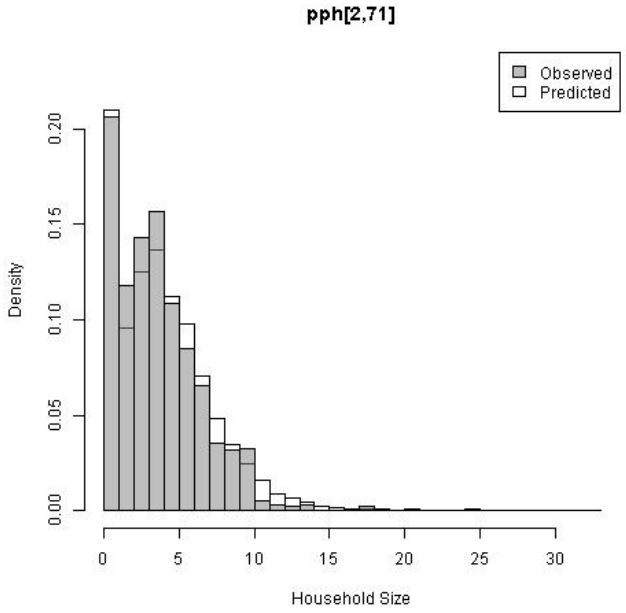

pph[2,67]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

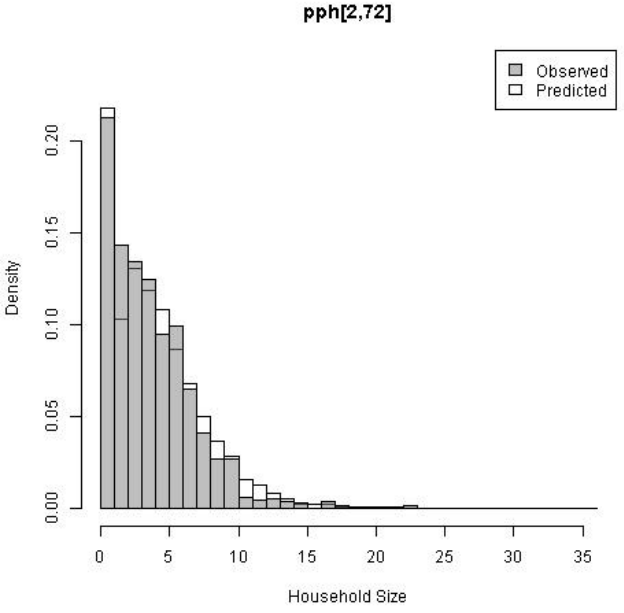

pph[2,68]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
Household Size

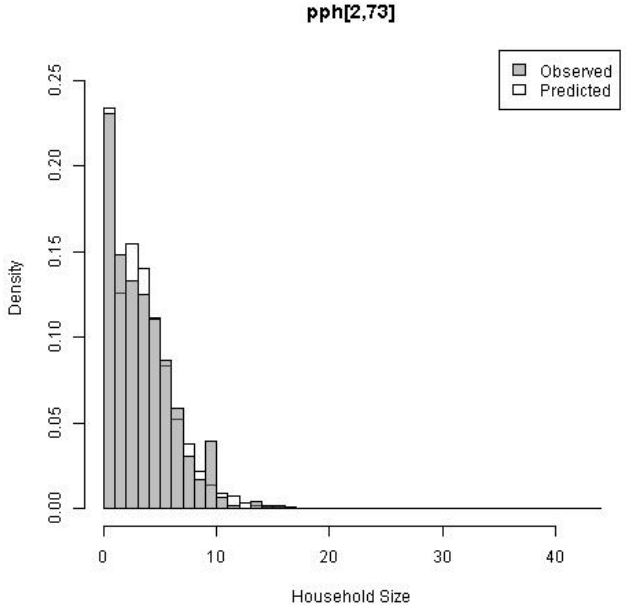

pph[2,69]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

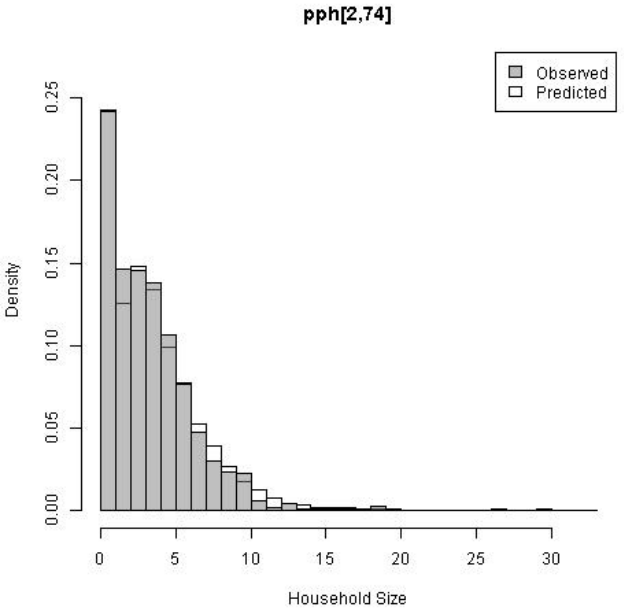

pph[2,7]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Household Size

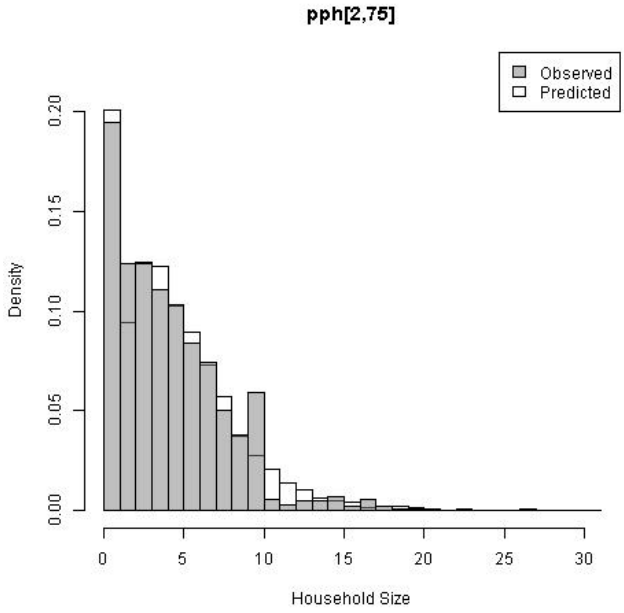

pph[2,70]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

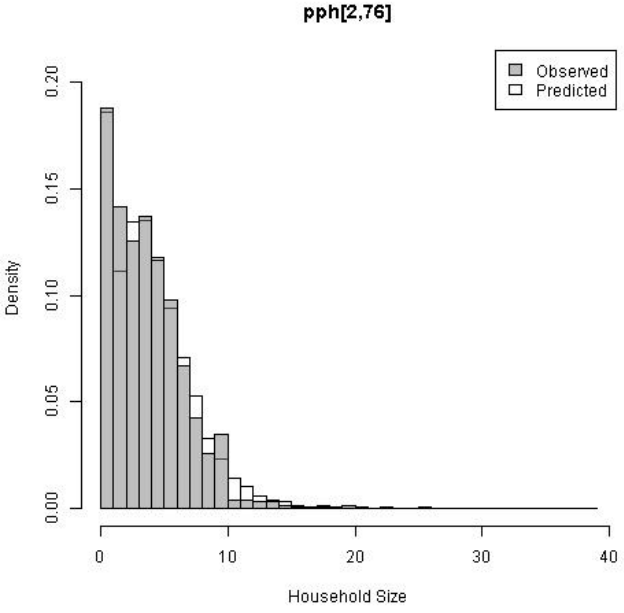

pph[2,71]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

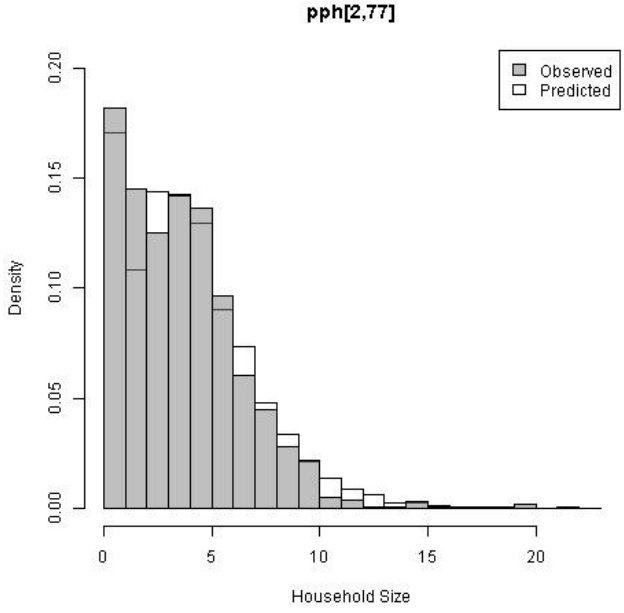

pph[2,72]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
35
Household Size

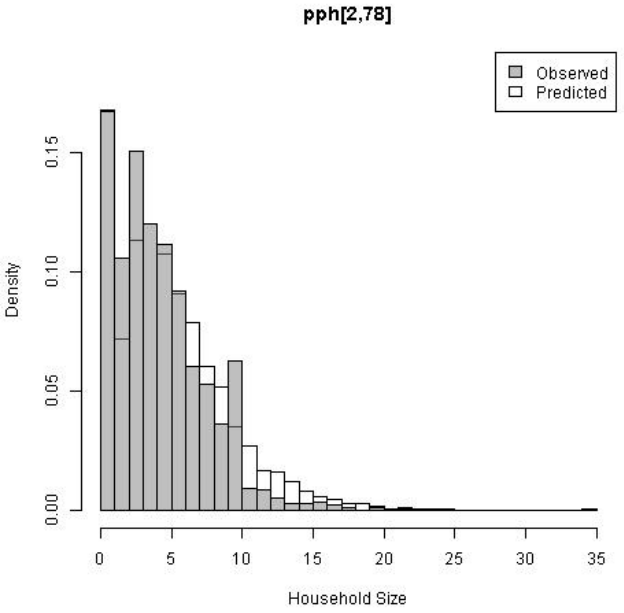

pph[2,73]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Household Size

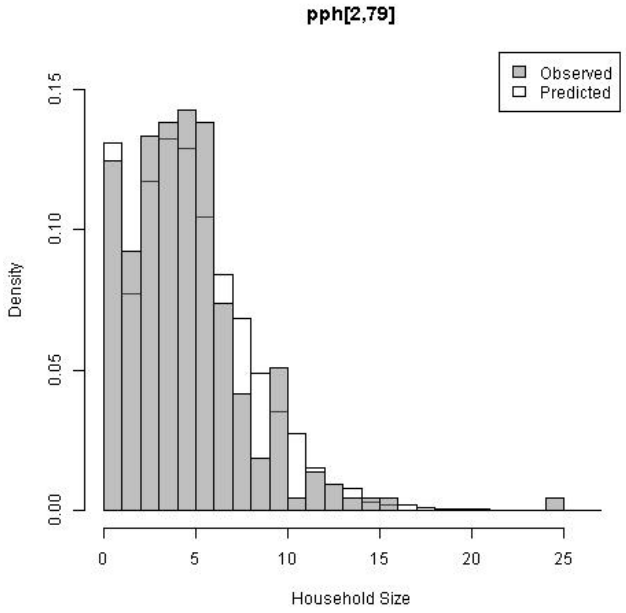

pph[2,74]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
Household Size

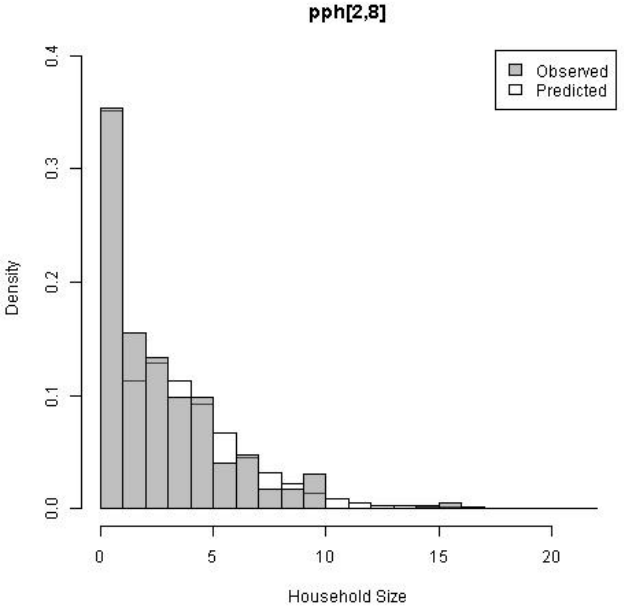

pph[2,75]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size

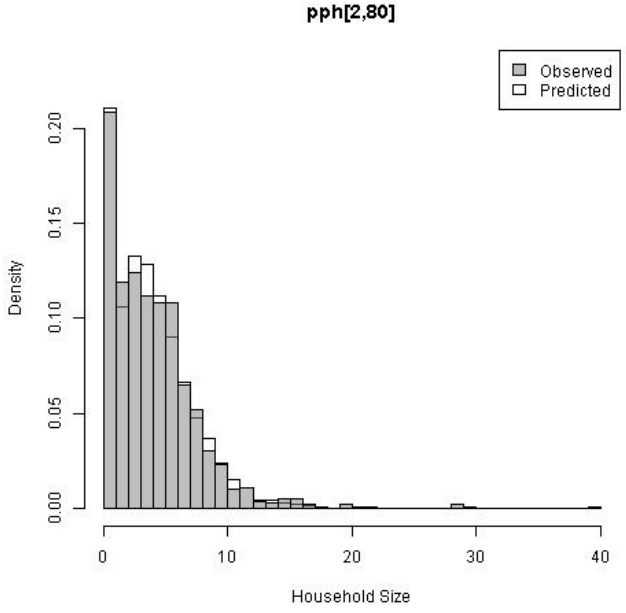

pph[2,76]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
Household Size

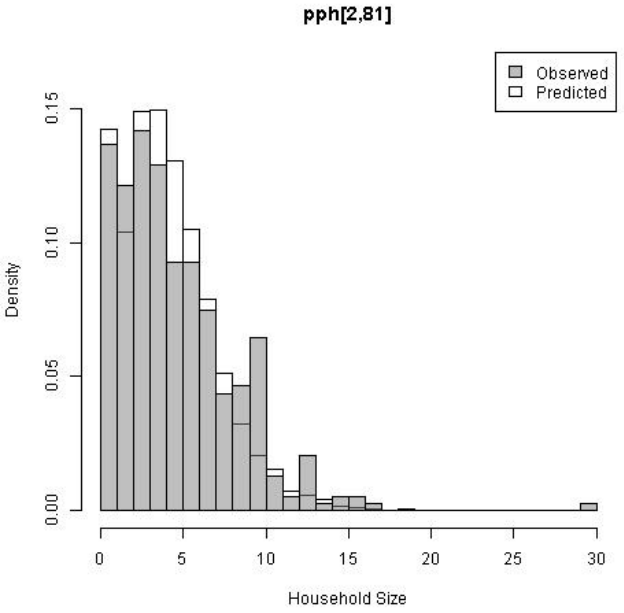

pph[2,77]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size

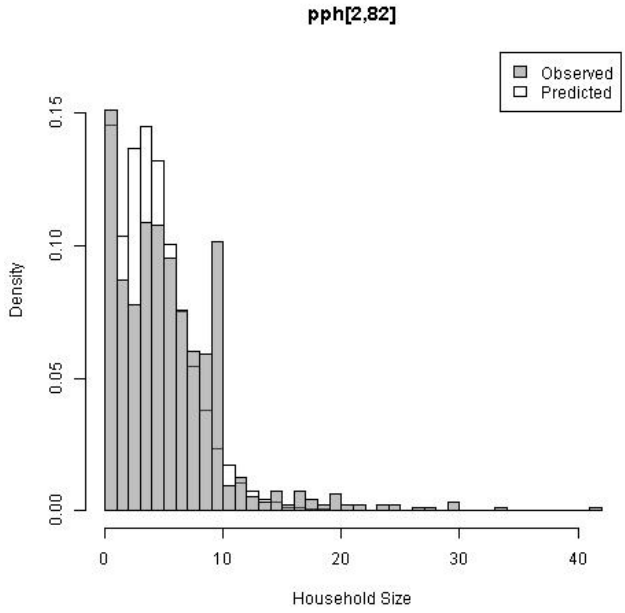

pph[2,78]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
35
Household Size

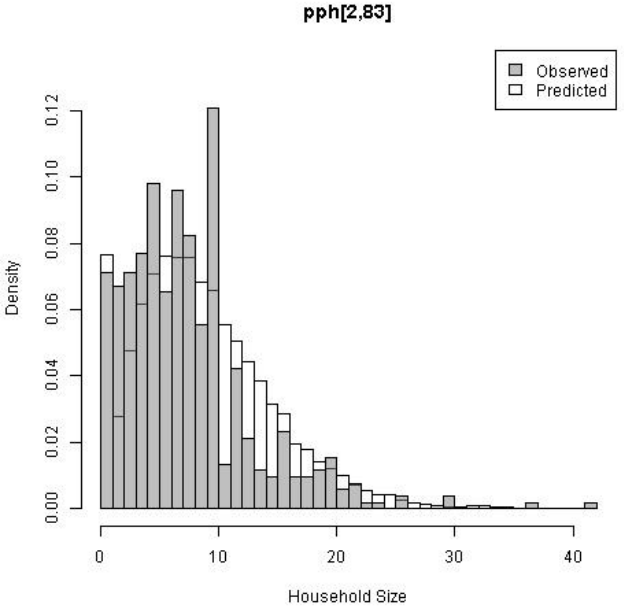

pph[2,79]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
Household Size

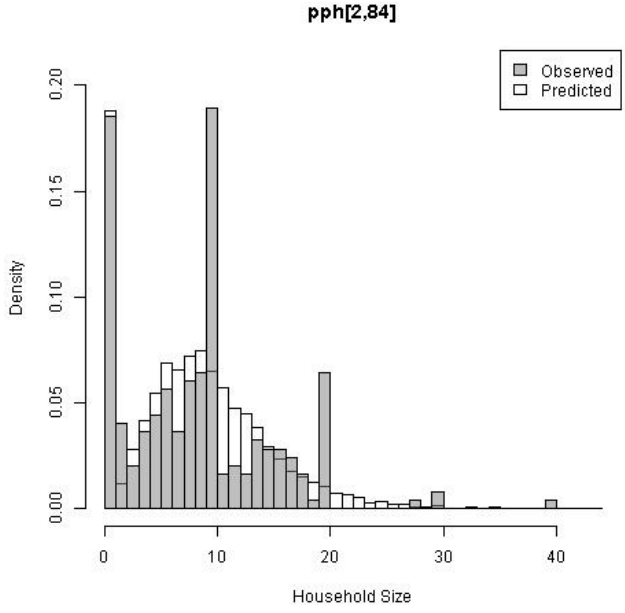

pph[2,8]
Observed
Predicted
Density
0.0
0.1
0.2
0.3
0.4
0
5
10
15
20
Household Size

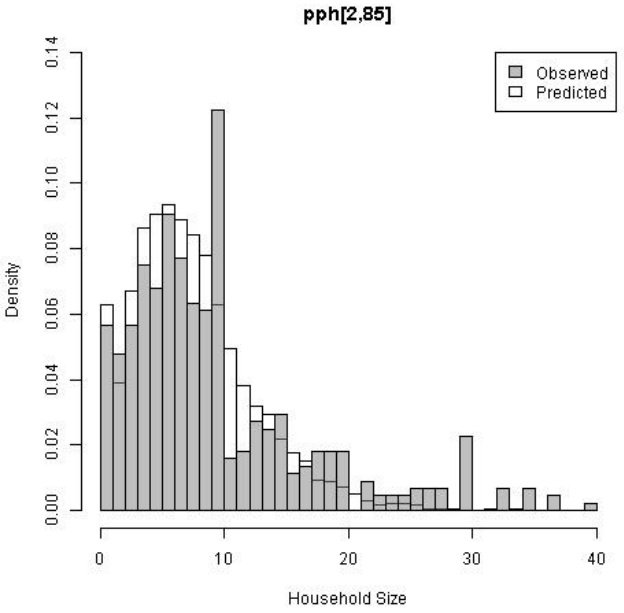

pph[2,80]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
Household Size

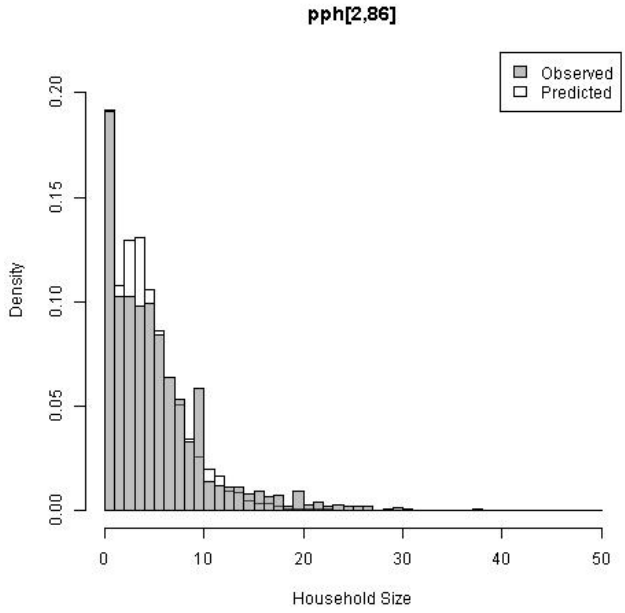

pph[2,81]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

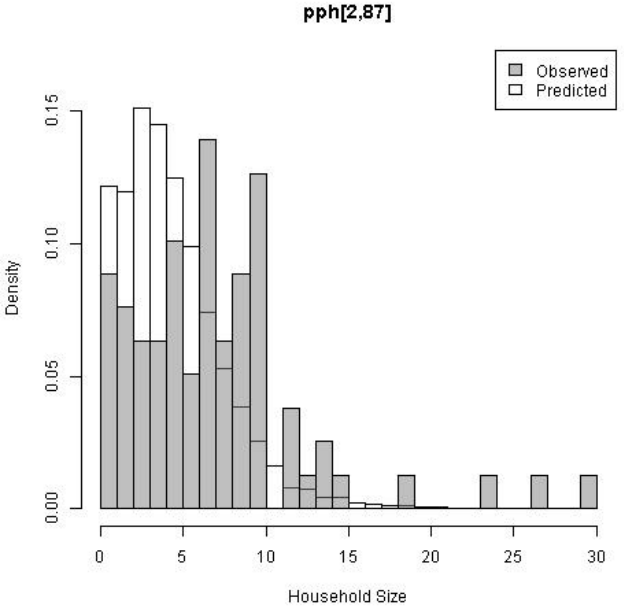

pph[2,82]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size

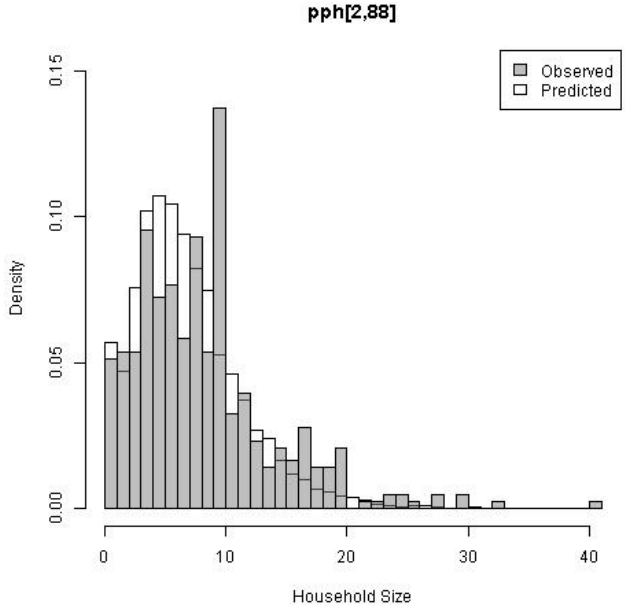

pph[2,83]
Observed
Predicted
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0
10
20
30
40
Household Size

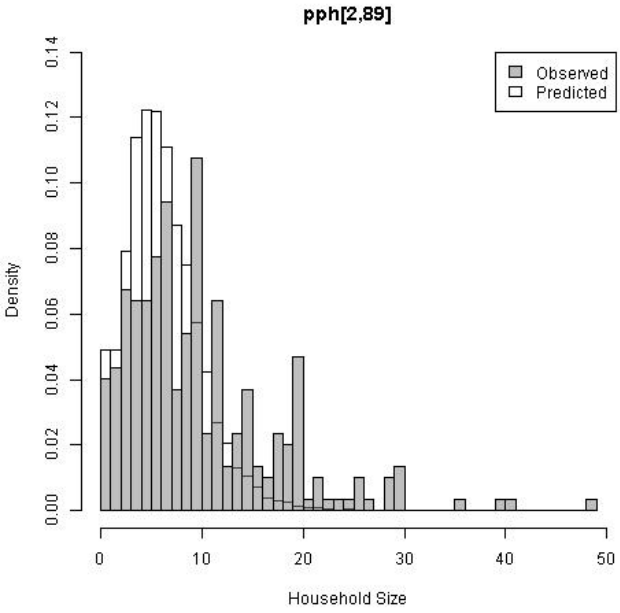

pph[2,84]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
Household Size

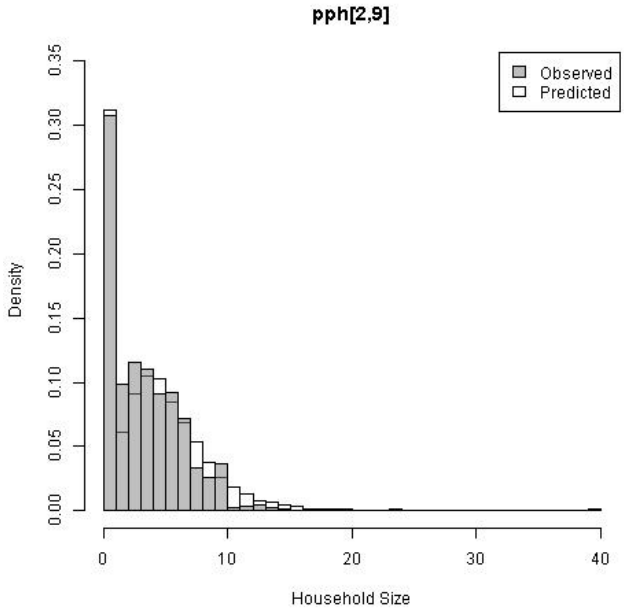

pph[2,85]
Observed
Predicted
Density
Household Size
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
10
20
30
40

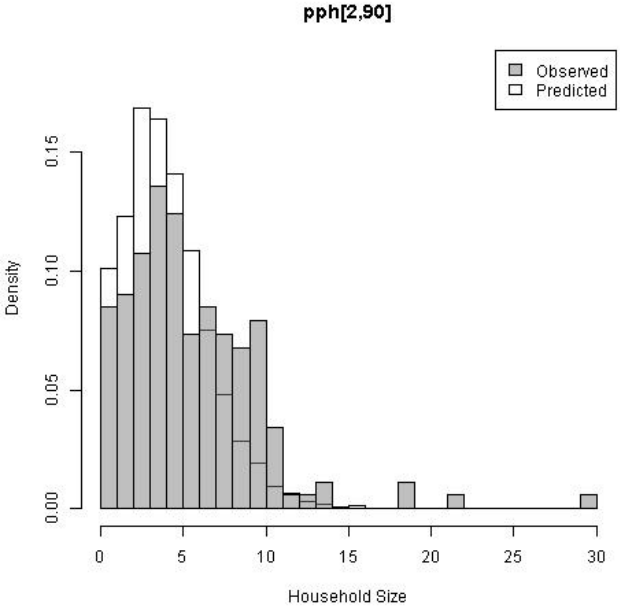

pph[2,86]
Observed
Predicted
Density
Household Size
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
50

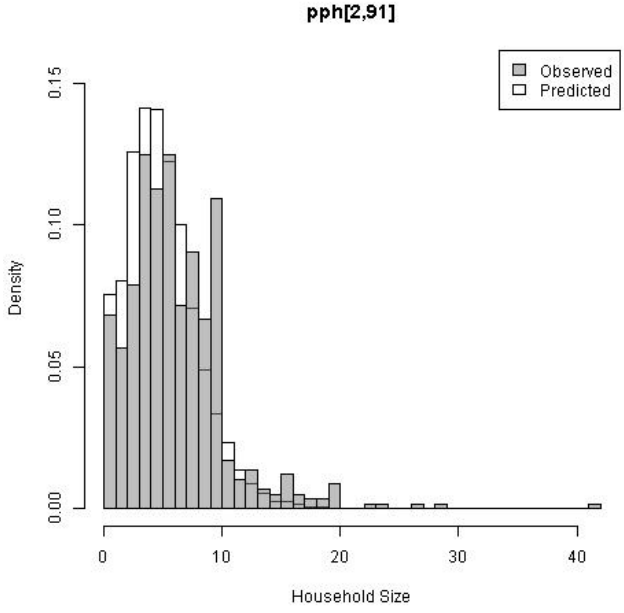

pph[2,87]
Observed
Predicted
Density
Household Size
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30

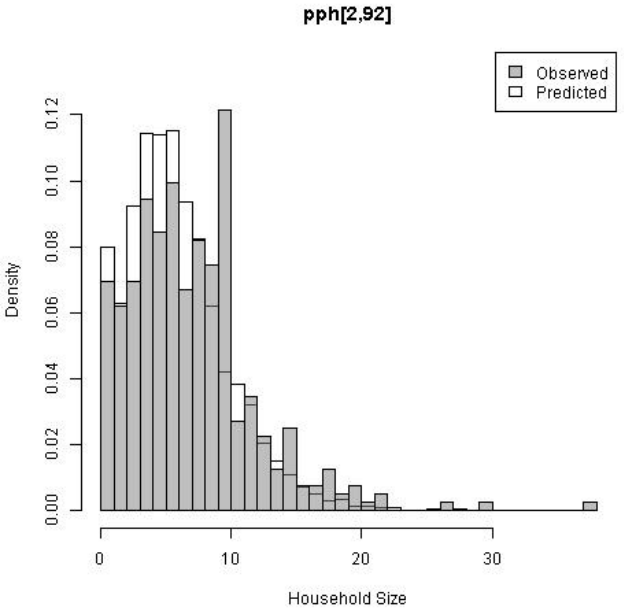

pph[2,88]
Observed
Predicted
Density
Household Size
0.00
0.05
0.10
0.15
0
10
20
30
40

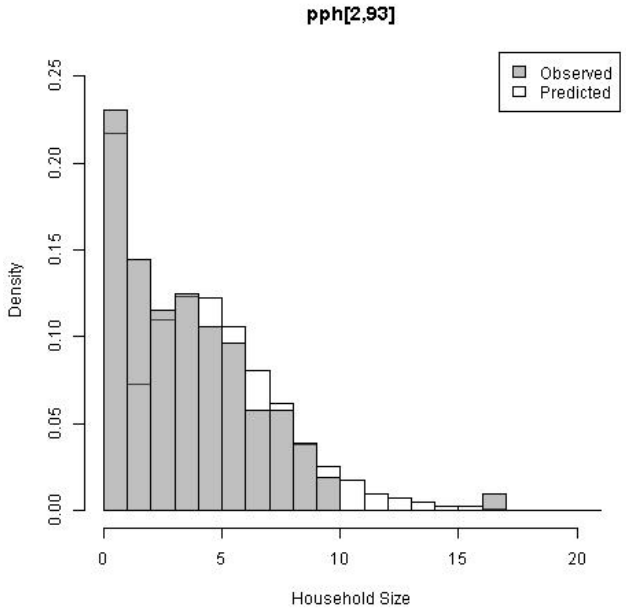

pph[2,89]
Observed
Predicted
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
10
20
30
40
50
Household Size

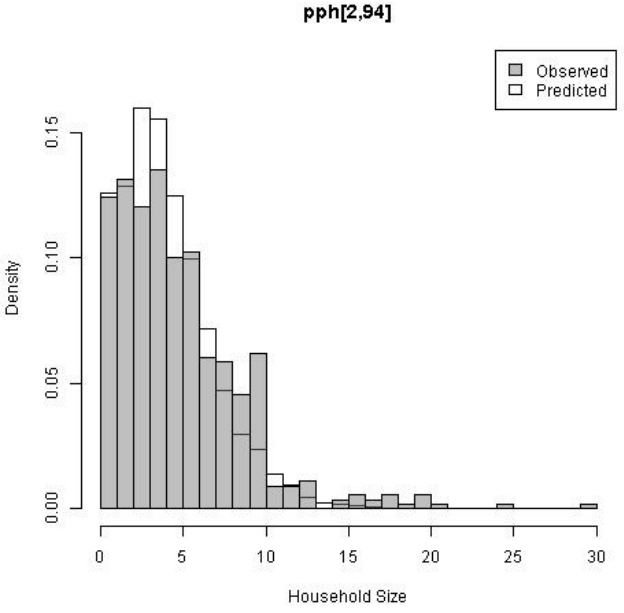

pph[2,9]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
Household Size

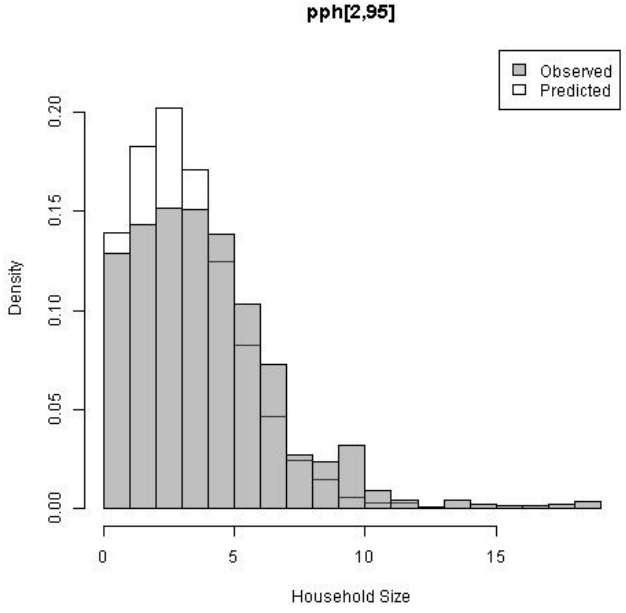

pph[2,90]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size

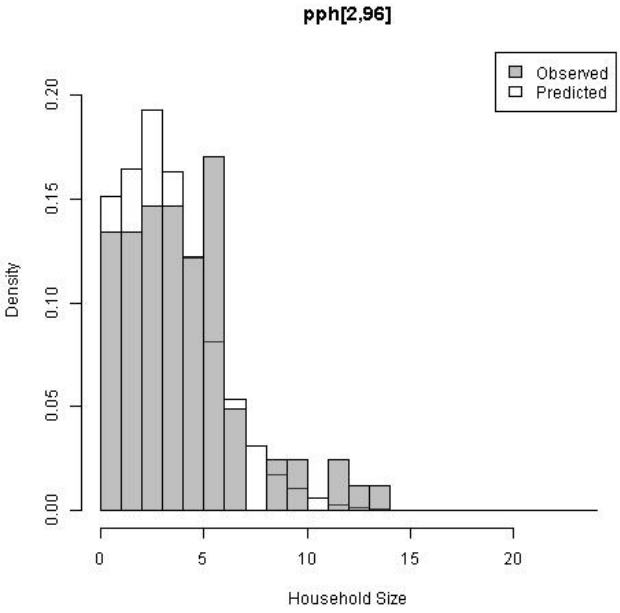

pph[2,91]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size

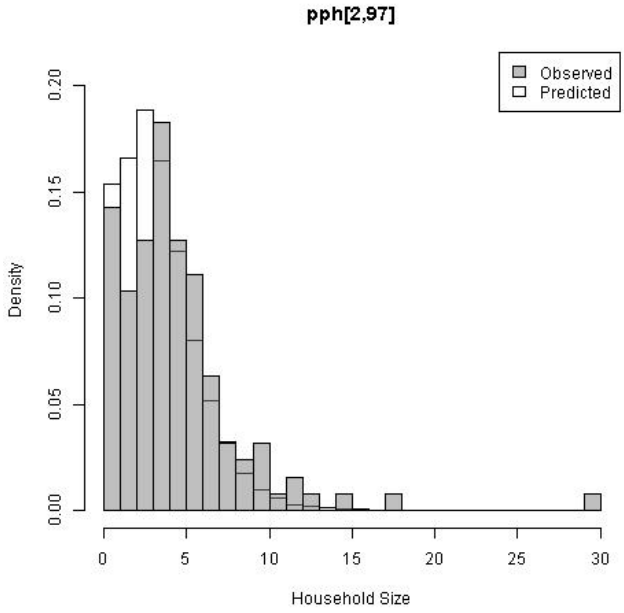

pph[2,92]
Observed
Predicted
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0
10
20
30
Household Size

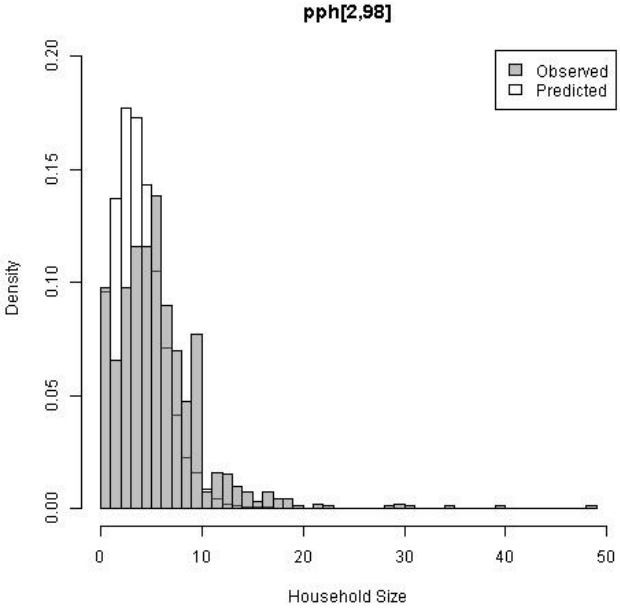

pph[2,93]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
Household Size

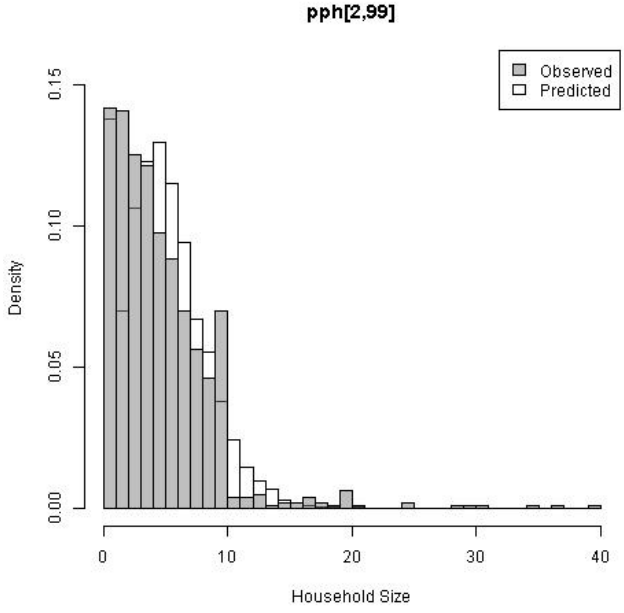

pph[2,94]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Household Size


pph[2,95]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
Household Size


pph[2,96]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Household Size


pph[2,97]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Household Size


pph[2,98]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
50
Household Size


pph[2,99]
Observed
Predicted
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Household Size