

## Minnesota Radon Levels

The data were collected by the EPA and the Minnesota Department of Health in November, 1987, for 1003 households in Minnesota. Radon concentrations were monitored in each house for two days using a charcoal canister. To maintain anonymity, houses were identified by county only. For each house in the survey, the county identification number and the two-day charcoal canister measurement of the radon level are available. Houses were selected county by county for the sample. Within a county, each house had an equal chance of being included in the survey.

Variable	Description
County ID	Identifier for the county in which the house is located.
Radon	Radon measurement, in picoCuries per liter (pCi/l).

The log-normal distribution is frequently used in environmental statistics. If the log-normal distribution fits the data reasonably well, then it can be used to provide estimates of the proportion of houses that exceed the recommended EPA action level, i.e. the level at which the home owner should remediate.

- Examine the data from a heavily sampled county, such as Hennepin, to determine how well the data fit the log-normal distribution. To do this, make a normal quantile plot to the log of the radon measurements.
- The data are truncated on the left due to limitations of the measurement process. That is, measurements below 0.5 pCi/l are reported as 0.5 pCi/l. Conduct a simulation study to determine how to adjust the estimates of the mean and SD to compensate for this truncation.
- According to the fitted density, what proportion of the houses in Hennepin County exceed 4 pCi/l? Provide a confidence interval for this proportion. To do this you will need to estimate the parameters of the log-normal and use these to estimate the probability of exceeding 4pCi/l. Then construct a confidence interval for this estimate.
- How does the above procedure differ if the average and SD of the log-normal distribution are estimated directly from the sampled radon measurements, as opposed to estimating it indirectly through the average and SD of the logged measurements? Conduct a simulation study that compares the properties of these estimates under various values of  $\mu$ ,  $\sigma$ , and  $n$ .