ADVANCED BAYESIAN MODELING

ROBUST INFERENCE: **THE *t* DISTRIBUTION**

Sometimes the data suggest that a stochastic value in a model is unusual, relative to what is typical for the model.

The stochastic value could be a data value y_i, or it could be a parameter θ_j in a hierarchical model.

In either case, it could be called an **outlier**.

Outliers occur for various reasons, including mistakes in the data, unused explanatory variables, and mis-specified distributions.

Typically, outliers in a quantitative variable are extreme (very large or small) compared to what the posterior would predict.

Models that use the normal distribution are especially susceptible to outliers because the normal density has **short (light) tails**:

k	Prob. a $\mathrm{N}(\mu, \sigma^2)$ variable is $> k\sigma$ from μ
1	0.3173105
2	0.0455003
3	0.0026998
4	0.0000633
5	0.0000006

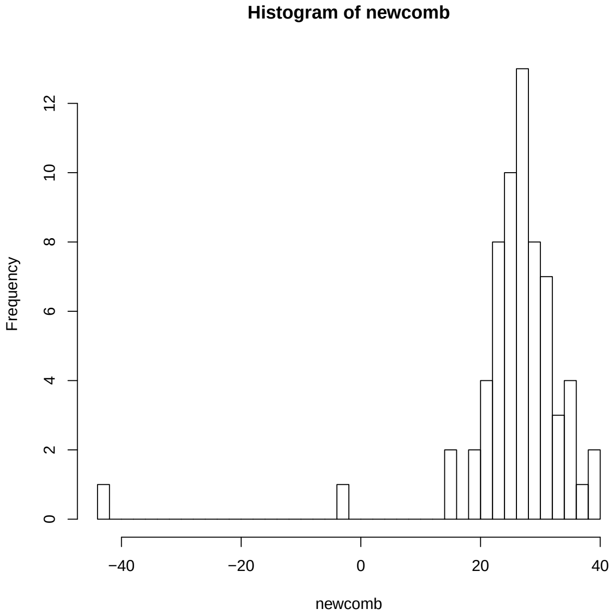
Newcomb Data

The `newcomb` data set comes from an early (1882) experiment to determine the speed of light. The 66 measurements are shifted and scaled times for light to travel the same known distance (in air).

Continuous measurements like these would typically be modeled with a normal distribution.

```
> library(MASS)  # has newcomb data set

> hist(newcomb, breaks=50)
```


Histogram of newcomb
Frequency
0
2
4
6
8
10
12
−40
−20
0
20
40
newcomb

Detection

The `newcomb` data set has obvious outliers, relative to what would be expected under a normal distribution.

In situations not as obvious, posterior predictive checks can be used (e.g., based on max and min statistics, or max and min of standardized errors).

BDA3, Sec. 6.3, has an example with the `newcomb` data.

To identify which particular observations are problematic, marginal predictive checks can be useful (BDA3, Sec. 6.3).

Robustness

For models with standard distributions (like the normal), outliers can be highly influential – removing them disproportionately affects the inference.

We seek models that are more **robust**: less sensitive to extreme values of a few observations (or parameters).

A robust model can be used as an alternative to a standard model, either in sensitivity analysis, or for improved inference.