# Teaching Statistics with Chocolate Chip Cookies

**Herbie Lee**

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

http://www.ams.ucsc.edu/~herbie

## *Learning Goals — Motivation*

- Motivate students to learn
- Get their attention
- Make them realize how statistics is relevant to their lives
- Re-assure them that statistics doesn't have to be scary
- Particularly important in a large lecture class
- Long-term recurring example

## *Learning Goals — Topics*

- Variability
- Inter-rater agreement and measurement error
- Exploratory data analysis: displays of data, outliers
- Poisson distribution, empirical distributions, extreme values
- Sampling distributions
- Hypothesis testing: one sample $t$-tests, two sample $t$-tests, and analysis of variance
- Bayesian statistics: prior elicitation and prior sensitivity

## *Why Cookies?*

- Many traditional examples don't appeal to the whole class
- Can have gender bias in interest in examples such as sports
- Everyone likes and understands cookies
- Snack foods are well-documented examples (Dyck and Gee, 1998; Rossman and Chance, 1999; Richardson and Haller, 2002; Richardson et al., 2005)

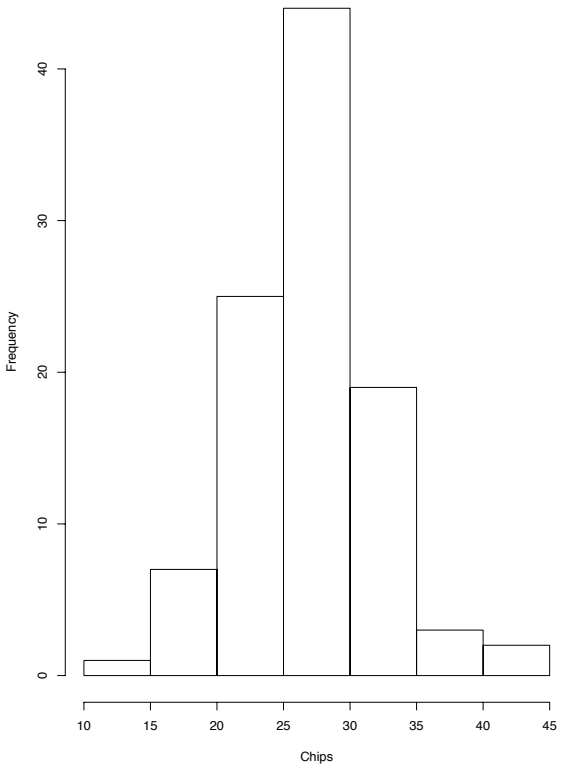
## *Counting Chips*

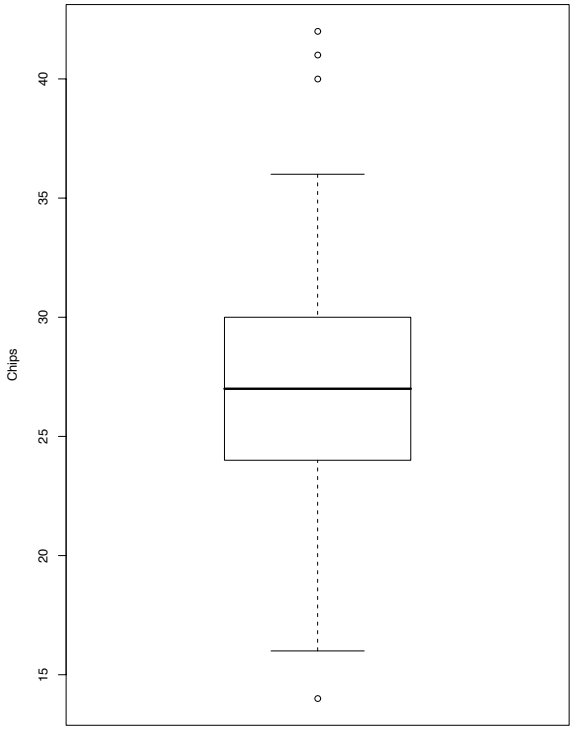
- Mass produced cookies (Chips Ahoy, Keebler, Store Brands)
- Each student counts their own
- Can't see all of the chips from the outside
- Report to whole class/collect data
- Discuss counting, inter-rater agreement, measurement error

## *Variability*

- Students are surprised by the variability
- Approximately Poisson
- "Striking Demonstrations" provoke excitement and learning (Sowey, 2001)

# *EDA and Graphical Displays*

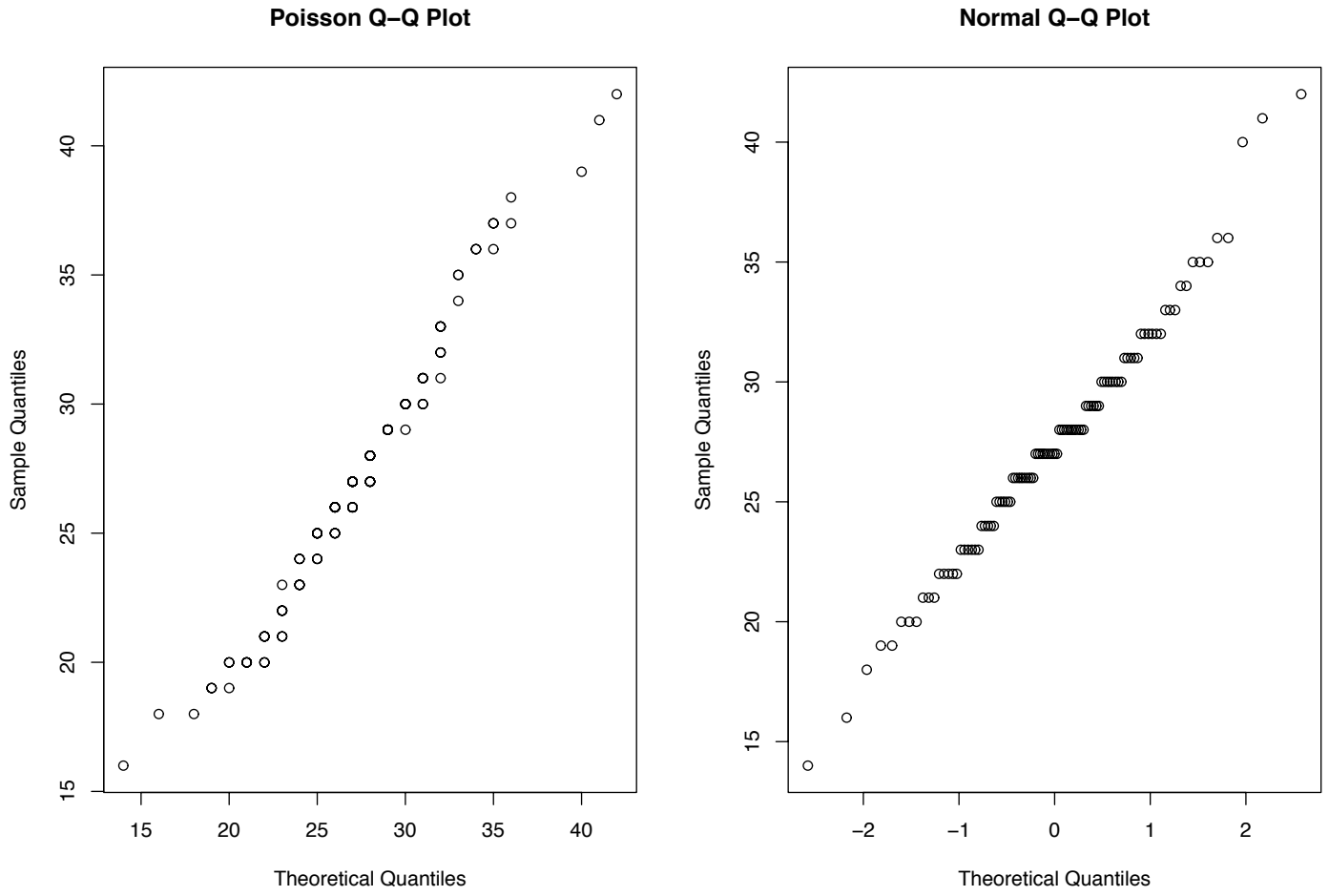


## *The Poisson Distribution*

- Can ask "If a cookie has 25 chips on average, what is the probability it has exactly 20 chips? At least 20 chips?"
- Histograms or Q-Q plots for comparing empirical and theoretical distributions
- Compare observed and expected variability
- Extreme values, maxima, minima

# Q-Q plots


Poisson Q–Q Plot
Sample Quantiles
Theoretical Quantiles
15
20
25
30
35
40
Normal Q–Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

## *Sampling Distributions*

- Can be difficult to understand the concept of the sampling distribution of the mean (Gourgey, 2000)
- Can look at average number of chips per bag
- Students understand the physical concept of the average over a bag

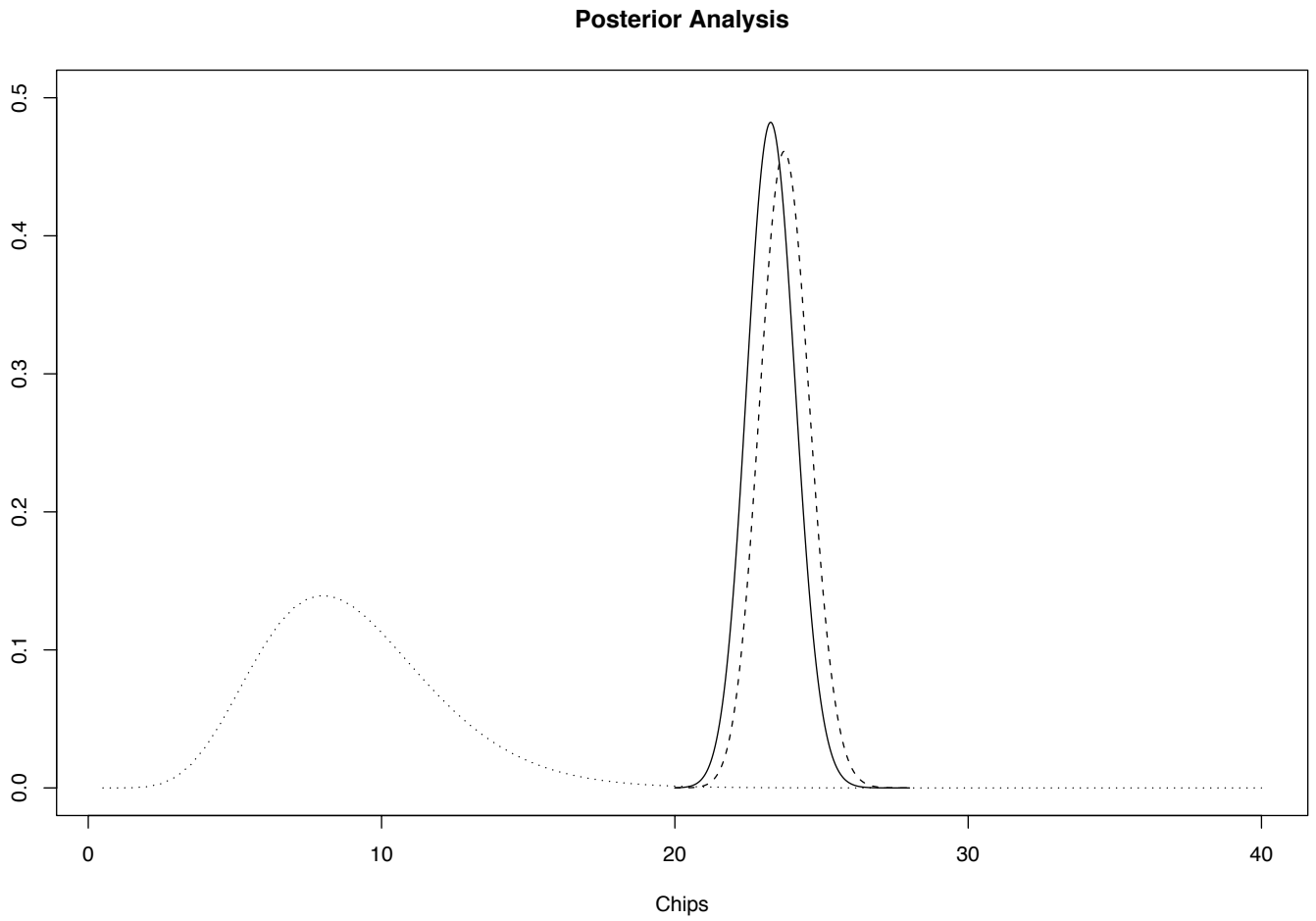
## *Hypothesis Tests*

- If the manufacturer claims an average of 27 chips per cookie, how would we decide if the manufacturer is truthful, using only a highly variable sample of cookies?
- Given how much difference there is in the counts for a single brand, how could we say if one brand of cookies has more chips than another?
- How could we compare more than two brands?

## *Bayesian Statistics*

- Prior elicitation
- Computation of posterior (conjugate setting)
- Posterior mean is weighted average of prior mean and data estimate
- Prior sensitivity

# *Prior, Likelihood, and Posterior*



## *Logistics*

- Can be done in a large or small class
- Need enough cookies for everyone
- Display results (e.g., laptop projection)
- Best done early in the term in an intro class
- Takes only 10-20 minutes

# References


Dyck, J. L. and Gee, N. A. (1998). "A Sweet Way to Teach Students About the Sampling Distribution of the Mean." *Teaching of Psychology*, 25, 3, 192–195.

Gourgey, A. F. (2000). "A Classroom Simulation Based on Political Polling To Help Students Understand Sampling Distributions." *Journal of Statistics Education*, 8, 3.

Lee, H. K. H. (2007). "Chocolate Chip Cookies as a Teaching Aid." *The American Statistician*, 61, 4, 351–355.

Richardson, M. and Haller, S. (2002). "What is the Probability of a Kiss? (It's Not What You Think)." *Journal of Statistics Education*, 10, 3.

Richardson, M., Rogness, N., and Gajewski, B. (2005). "4 out of 5 Students Surveyed Would Recommend this Activity (Comparing Chewing Gum Flavor Durations)." *Journal of Statistics Education*, 13, 3.

Rossman, A. J. and Chance, B. L. (1999). "Teaching the Reasoning of Statistical Inference: A 'Top Ten' List." *College Mathematics Journal*, 30, 4, 297–305.

Sowey, E. R. (2001). "Striking Demonstrations in Teaching Statistics." *Journal of Statistics Education*, 9, 1.