

Calculating statistics

Jeff Leek

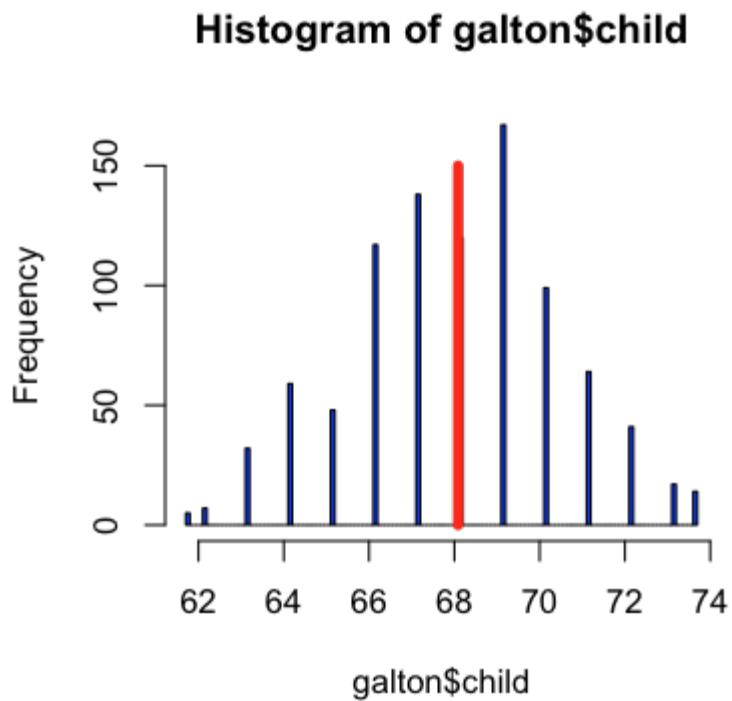
@jtleek

www.jtleek.com

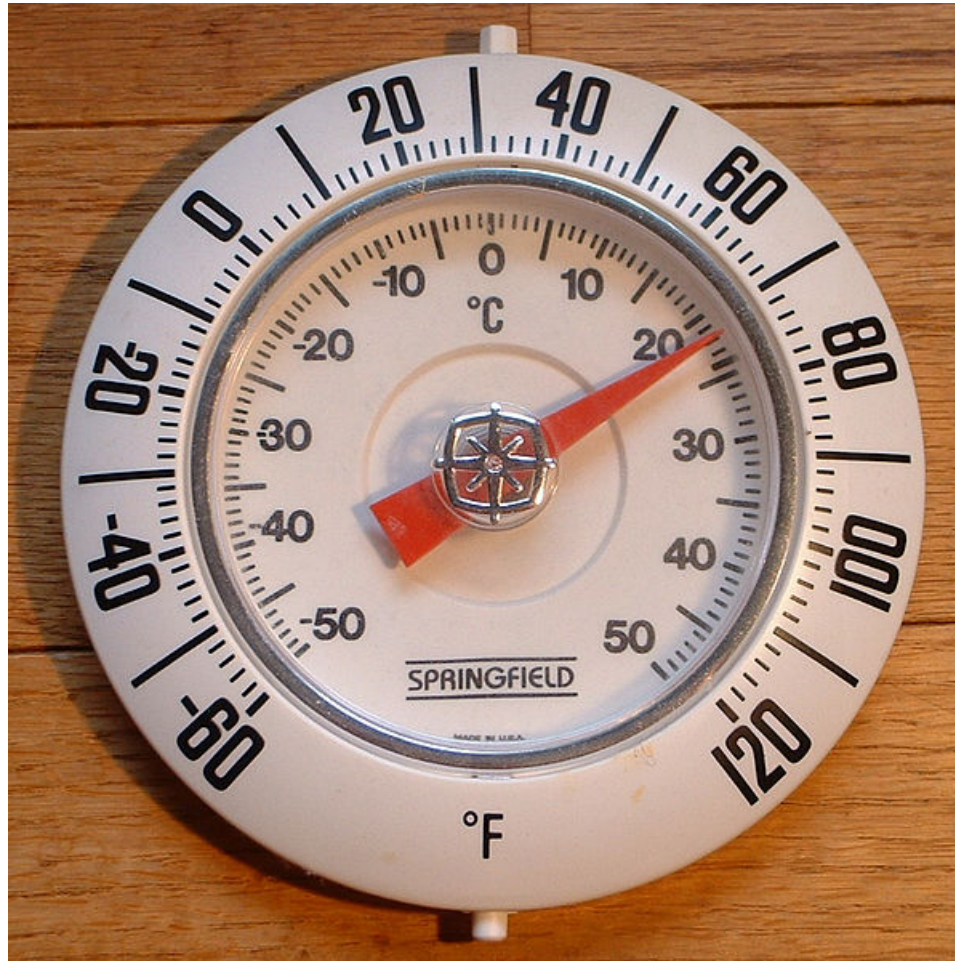
[OBJ] Typical goal: quantify how certain we are about associations

$$\bar{X} = \frac{1}{M} \sum_{i=1}^M X_i$$

$$\bar{X} = 70$$



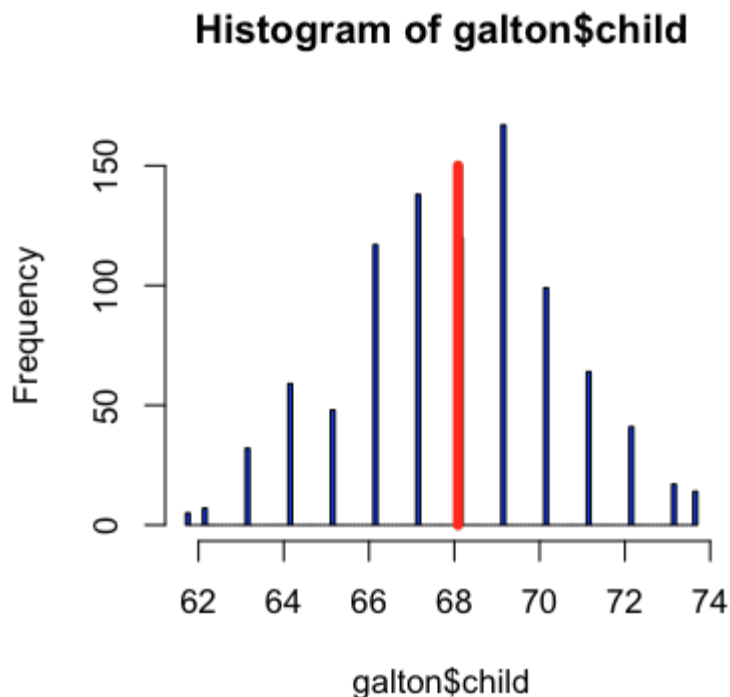
Standardization of distance is critical



$$s_X^2 = \frac{1}{M-1} \sum_{i=1}^M (X_i - \bar{X})^2$$

$$\bar{X} = \frac{1}{M} \sum_{i=1}^M X_i$$

$$\frac{\bar{X} - 70}{s_x / \sqrt{n}}$$



Observations:

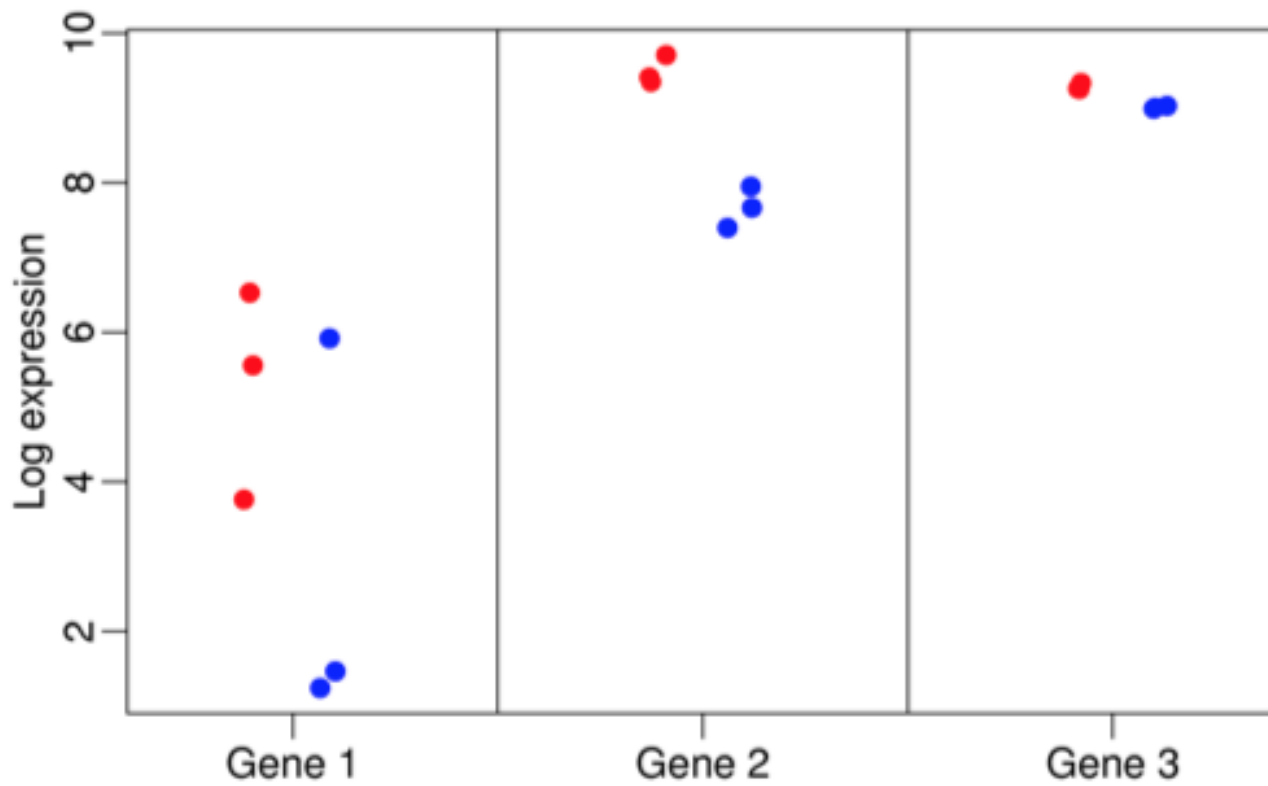
$$X_1, \dots, X_M \quad Y_1, \dots, Y_N$$

Averages:

$$\bar{X} = \frac{1}{M} \sum_{i=1}^M X_i \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

SD² or variances:

$$s_X^2 = \frac{1}{M-1} \sum_{i=1}^M (X_i - \bar{X})^2 \quad s_Y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$$



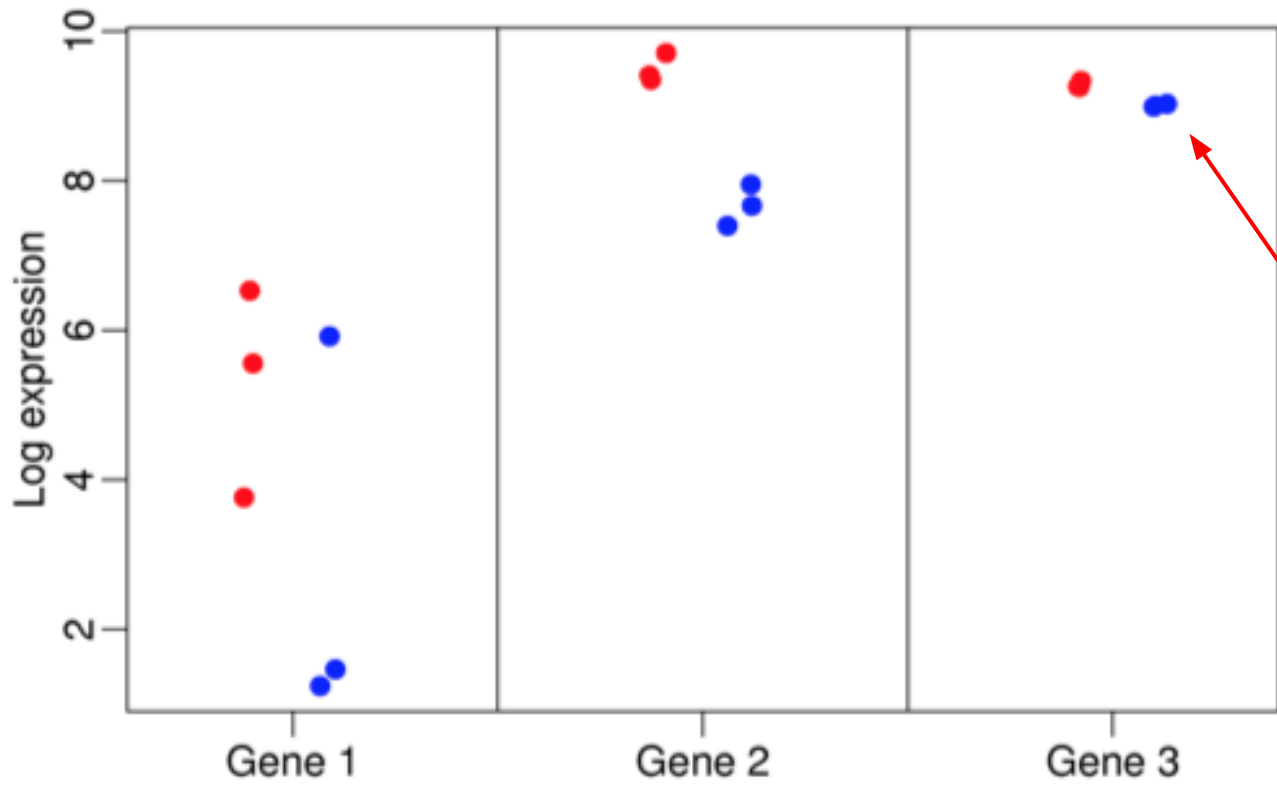
The most popular statistic
The t-statistic

$$\frac{\bar{Y} - \bar{X}}{\sqrt{\frac{s_Y^2}{N} + \frac{s_X^2}{M}}}$$

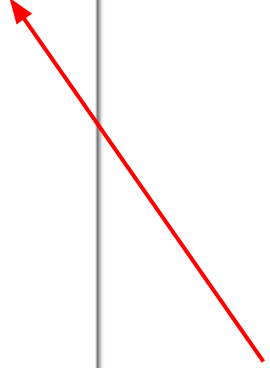
**T-statistics for linear models
Called Wald statistics**


$$\text{Expr} = b_0 + b_1 \text{Age} + e$$

$$\frac{\hat{b}}{s.e.(\hat{b})}$$



This happens a lot



Moderated statistics

Remove bias due to small variability

$$\text{Expr} = b_0 + b_1 \text{Age} + e$$

$$\frac{\hat{b}}{s.e.(\hat{b})} + c$$

Notes and further reading

- Linear models for microarray data
 - <http://www.statsci.org/smyth/pubs/limma-biocbook-reprint.pdf>
- Statistics and R for the Life Sciences
 - <https://www.edx.org/course/statistics-r-life-sciences-harvardx-ph525-1x>