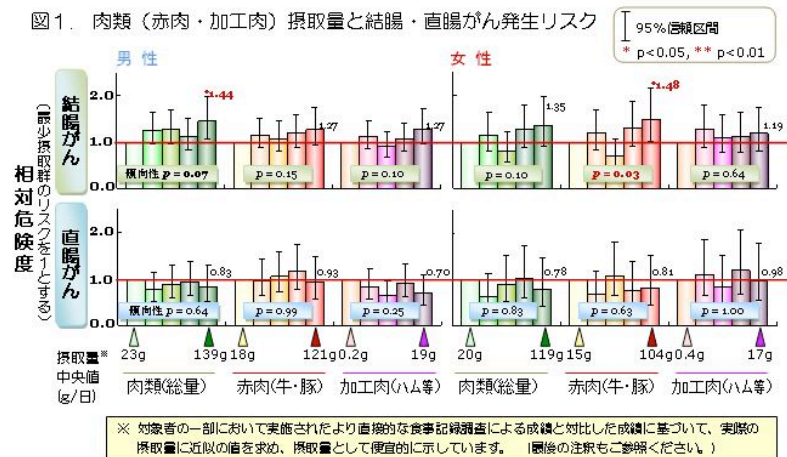# Statistics/Data Analysis

・ Statistics is decriptive or confirmatory

  - acting like a judge with what's available

  - testing specific theories

  - not suitable for observational or large datasets

・ Exploratory data analysis is open-ended

  - acting like a detective

  - generating new theories

From models to graphics (1)

- t test (t.test)

- Binomial test (binom.test, prop.test)

- Chisq test (chisq.test)

- Nonparametric tests (wilcox.test)

- Correlation (cor, cor.test)

- Regression, anova and linear models (lm, anova)

From models to graphics (2)

- glm (including loglinear models, logistic regression)

- Density estimation (density, ks, KernSmooth …)

- Kolmogorov Smirnov (ks.test)

- Outliers (extremevalues, mvoutlier, rainbow …)

- Bootstrapping (boot)

- Survival analysis (survival, party…)

- Cluster analysis (hclust, agnes, …)

- …

See discussion at:
andrewgelman.com/2011/12/looking-at-many-comparisons-may-increase-the-risk-of-finding-something-statistically-significant-by-epidemiologists-a-population-with-relatively-low-multilevel-modeling-consumption/