UF UNIVERSITY of FLORIDA

Statistical Hypothesis: χ^2 test and Mendelian segregation distortion

Postdoc: Luís Felipe Ventorim Ferrão
Professor: Patricio Munoz

September 6, 2017

Why test statistical hypothesis?

- Hypothesis testing is an essential procedure in statistics.
- A hypothesis test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

In the linkage mapping context, statistical test are useful:

- Verification of Mendelian segregation distortion (χ^2 tests).
- Linkage analysis of markers (odds ratios, i.e. the ratio of linkage versus no linkage).

Step by step: Hypothesis Testing

Cartoon guide to statistics

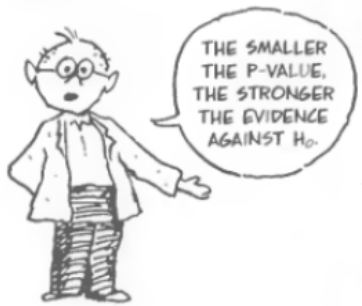
Step 1. *FORMULATE ALL HYPOTHESES.*

H_0, THE *NULL HYPOTHESIS*, IS USUALLY THAT THE OBSERVATIONS ARE THE RESULT PURELY OF *CHANCE*.

H_a, THE *ALTERNATE HYPOTHESIS*, IS THAT THERE IS A REAL EFFECT, THAT THE OBSERVATIONS ARE THE RESULT OF THIS REAL EFFECT, PLUS CHANCE VARIATION.

Step 2. THE *TEST STATISTIC*. IDENTIFY A STATISTIC THAT WILL ASSESS THE EVIDENCE AGAINST THE NULL HYPOTHESIS.

Step 3. *P-VALUE:* A PROBABILITY STATEMENT WHICH ANSWERS THE QUESTION: IF THE NULL HYPOTHESIS WERE TRUE, THEN WHAT IS THE PROBABILITY OF OBSERVING A TEST STATISTIC AT LEAST AS EXTREME AS THE ONE WE OBSERVED?



Step 4. COMPARE THE P-VALUE TO A FIXED *SIGNIFICANCE LEVEL*, α.

α ACTS AS A CUT-OFF POINT BELOW WHICH WE AGREE THAT AN EFFECT IS STATISTICALLY SIGNIFICANT. THAT IS, IF

$$\text{P-VALUE} \leq \alpha$$

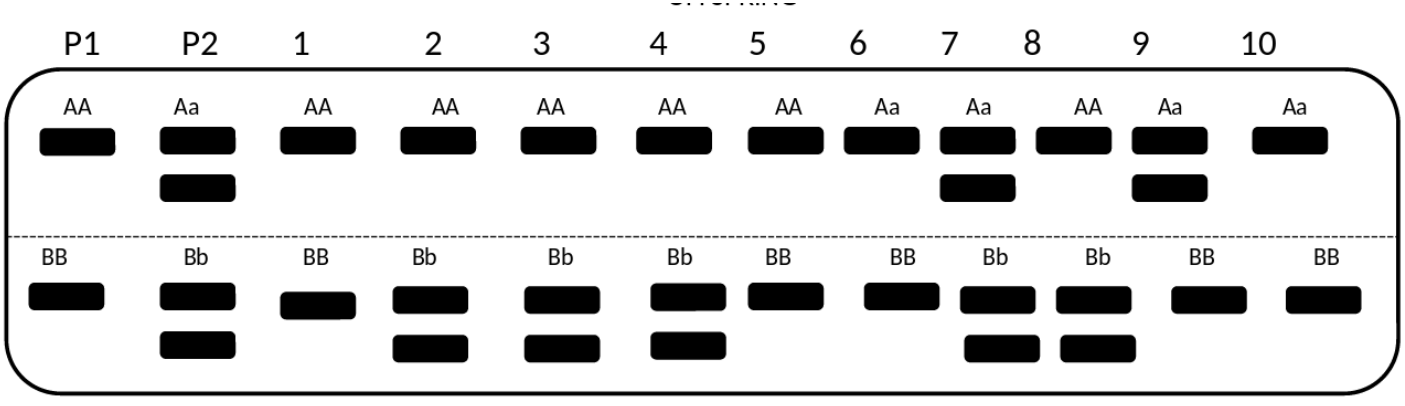
THEN WE *RULE OUT THE NULL HYPOTHESIS* H_0 AND AGREE THAT SOMETHING ELSE IS GOING ON.

Mendelian segregation distortion

- The construction of a linkage map requires a segregating population.
- There are expected segregation ratios in accordance to the population type and marker system (Codominant or Dominant).
- Significant deviations from expected ratios can be analysed using chi-square tests.
- Markers with segregation distortion can be caused by biological reasons (eg, deleterious genes and chromosomal rearrangements) or sample problems (eg, population size).
- Inclusion of distorted markers can be problematic during the estimation of the recombination rate.

Mendelian segregation distortion
Practical example

- Backcross, 2 markers (A and B) and 10 full-sib individuals.



P1 = Progenitor 1
P2 = Progenitor 2

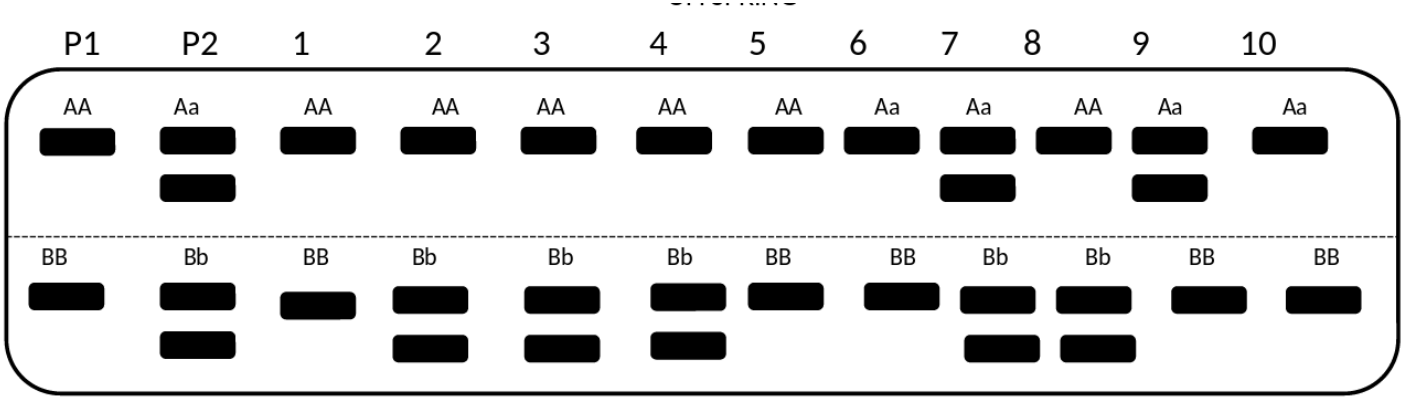
Marker	Expected Segregation	Observed Segregation
A	5 (AA) | 5 (Aa) (1:1)	8 (AA) | 2 (Aa)
B	5 (BB) | 5 (Bb) (1:1)	5 (BB) | 5 (Bb)

Questions:

- Intuitively: which marker showed a Mendelian segregation distortion?
- Formally: how to test for the segregation distortion?

Mendelian segregation distortion
Practical example

- Backcross, 2 markers (A and B) and 10 full-sib individuals.



P1 = Progenitor 1
P2 = Progenitor 2

Marker	Expected Segregation	Observed Segregation
A	5 (AA) | 5 (Aa) (1:1)	8 (AA) | 2 (Aa)
B	5 (BB) | 5 (Bb) (1:1)	5 (BB) | 5 (Bb)

Questions:

- Intuitively: which marker showed a Mendelian segregation distortion?
- Formally: how to test for the segregation distortion?

Mendelian segregation distortion

χ^2 test - Backcross

Hypotheses formulation

- H0 (null hypothesis): the data are consistent with a specified distribution (1:1 segregation).
- Ha (alternative hypothesis): the data are not consistent with a specified distribution.

Genotype	AA	aa
Exp. freq	1/2	1/2
n. exp	n/2	n/2
n. obs	n_1	n_2

$$\chi^2 = \sum \frac{(n.obs - n.exp)^2}{n.exp} = \frac{(n_1 - n/2)^2}{n/2} + \frac{(n_2 - n/2)^2}{n/2}$$

$$= \frac{(n_1 - n_2)^2}{n} \sim \chi^2_{1df}$$

Mendelian segregation distortion

χ^2 test - Backing to the example

Marker	Marker A		Marker B	
Exp. Freq	1/2	1/2	1/2	1/2
Exp.n	5 (AA)	5 (Aa)	5 (BB)	5 (Bb)
Obs.n	5	5	8	2
χ^2	0		3.6	
p-value	1		0.05	

Last Step: Compare the p.value to a fixed significant level ($\alpha = 0.1$, for example)

- Marker A (p.value $> \alpha$) fail to reject the null hypothesis (No segregation distortion)
- Marker B (p.value $< \alpha$): reject the null hypothesis (Segregation distortion)
- Tip: check the chisq.test function in R, to performs chi-squared contingency table tests and goodness-of-fit tests.

Mendelian segregation distortion

Important points during the analysis of real data sets

- Multiple test correction (eg: Bonferroni, FDR and others)
- There is no consensus in the literature if distorted markers should be discarded.
- Important evolutionary events may explain distortion in the segregation pattern.
- Full-sib families derived from two outbred parents: mixed set of different marker types containing various segregation patterns, such as 1:1:1:1, 1:2:1, 3:1 and 1:1.