Dad phase unknown

odds ratio = \frac{1/2[(1-r)^n \times r^k] + 1/2[(1-r)^n \times r^k]}{0.5 \times \text{total # meioses}}

What single r value best explains the data?

For this, you need to search r’s.

Oops: a numerical mistake (thanks to Jonathan for detective work)

In real life this correction does matter…

Using only one phase

Accounting for both phases
Locus heterogeneity

age of onset

Table 1. Lod scores for linkage of breast cancer to 17q21/chr17:21. For each family, it is the mean age of diagnosis of breast cancer.

Recombination fraction

<table>
<thead>
<tr>
<th>Lod score</th>
<th>0.00</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lod score</td>
<td>2.02</td>
<td>1.56</td>
<td>1.08</td>
<td>0.60</td>
<td>0.18</td>
<td>-0.26</td>
</tr>
<tr>
<td>Lod score</td>
<td>2.75</td>
<td>2.00</td>
<td>1.31</td>
<td>0.90</td>
<td>0.50</td>
<td>0.10</td>
</tr>
<tr>
<td>Lod score</td>
<td>3.37</td>
<td>2.50</td>
<td>1.49</td>
<td>0.94</td>
<td>0.44</td>
<td>0.06</td>
</tr>
<tr>
<td>Lod score</td>
<td>3.97</td>
<td>2.85</td>
<td>1.71</td>
<td>1.04</td>
<td>0.60</td>
<td>0.18</td>
</tr>
<tr>
<td>Lod score</td>
<td>4.53</td>
<td>3.10</td>
<td>2.00</td>
<td>1.49</td>
<td>0.80</td>
<td>0.24</td>
</tr>
<tr>
<td>Lod score</td>
<td>5.05</td>
<td>3.35</td>
<td>2.30</td>
<td>1.71</td>
<td>1.04</td>
<td>0.60</td>
</tr>
<tr>
<td>Lod score</td>
<td>5.53</td>
<td>3.55</td>
<td>2.50</td>
<td>1.90</td>
<td>1.24</td>
<td>0.80</td>
</tr>
<tr>
<td>Lod score</td>
<td>6.01</td>
<td>3.75</td>
<td>2.70</td>
<td>2.10</td>
<td>1.49</td>
<td>0.80</td>
</tr>
<tr>
<td>Lod score</td>
<td>6.49</td>
<td>3.90</td>
<td>2.90</td>
<td>2.30</td>
<td>1.71</td>
<td>1.04</td>
</tr>
<tr>
<td>Lod score</td>
<td>6.97</td>
<td>4.00</td>
<td>3.10</td>
<td>2.50</td>
<td>1.90</td>
<td>1.24</td>
</tr>
<tr>
<td>Lod score</td>
<td>7.45</td>
<td>4.10</td>
<td>3.30</td>
<td>2.70</td>
<td>2.10</td>
<td>1.49</td>
</tr>
<tr>
<td>Lod score</td>
<td>7.93</td>
<td>4.20</td>
<td>3.50</td>
<td>2.90</td>
<td>2.30</td>
<td>1.71</td>
</tr>
<tr>
<td>Lod score</td>
<td>8.41</td>
<td>4.30</td>
<td>3.70</td>
<td>3.10</td>
<td>2.50</td>
<td>1.90</td>
</tr>
<tr>
<td>Lod score</td>
<td>8.89</td>
<td>4.40</td>
<td>3.90</td>
<td>3.30</td>
<td>2.70</td>
<td>2.10</td>
</tr>
<tr>
<td>Lod score</td>
<td>9.37</td>
<td>4.50</td>
<td>4.10</td>
<td>3.50</td>
<td>2.90</td>
<td>2.30</td>
</tr>
<tr>
<td>Lod score</td>
<td>9.85</td>
<td>4.60</td>
<td>4.30</td>
<td>3.70</td>
<td>3.10</td>
<td>2.50</td>
</tr>
<tr>
<td>Lod score</td>
<td>10.33</td>
<td>4.70</td>
<td>4.50</td>
<td>3.90</td>
<td>3.30</td>
<td>2.70</td>
</tr>
</tbody>
</table>

Odds = \frac{P(\text{your flips} | r)}{P(\text{your flips} | r = 0.5)} = \frac{(1-r)^n \cdot r^k}{0.5^{\text{total flips}}}

Odds ratio of model that coin is biased, relative to null

Table 2. Observed numbers of heads in two flips.

<table>
<thead>
<tr>
<th>Heads</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Significance cutoff

For single family testing, a significance level of p < 0.001 is typically used.
The analogy again

Testing lots of markers for linkage to a trait is analogous to having lots of students, each flipping a coin.

The search for the coin’s bias parameter is analogous to the search for recombination distance between markers and disease locus.

The analogy again

Testing lots of markers for linkage to a trait is analogous to having lots of students, each flipping a coin.

The search for the coin’s bias parameter is analogous to the search for recombination distance between markers and disease locus.

Each student is analogous to a marker.

Each coin flip is analogous to a family member in pedigree.

Multiple testing, shown another way

1. Simulate thousands of markers, inherited from parents to progeny.
2. Assign some family members to have a disease, others not.
3. Test for linkage between disease and markers, knowing there is none.

A real world scenario

You have invested a bolus of research money in a linkage mapping study of a genetic disease segregating in families. For each family member, you do genotyping at a bunch of markers.

When you finally run the linkage calculation, the strongest marker gives a LOD of 2. You desperately want to believe this is significant.

You simulate a fake trait with no genetic control 1000 times.

You find that in 433 of these simulations, the fake trait had a LOD > 2.

This means that in your real data, the probability of your precious linkage peak being a false positive is 433/1000 = 0.433.

If you spent more money and time to follow this up, it could be a complete waste. Essential to know.
Simulate 1000 times, ask how frequently you get a peak over a certain threshold.

With modest marker spacing in a human study, LOD of 3 is 9% likely to be a false positive.

But this would change in a different organism, with different number of markers, etc.

So in practice, everyone does their own simulation specific to their own study.
More markers = more tests = more chance for spurious high linkage score.

Not true when you add individuals (patients)! Always improves results.

Multiple testing in genetics

Marker density matters

But if the only marker you test is >50 cM away, will get no linkage.
Marker density matters

But if the only marker you test is >50 cM away, will get no linkage.

So a mapping experiment is a delicate balance between too much testing and not enough…

Candidate gene approach: apple pigment

Candidate gene approach:

Hypothesize that causal variant will be in known pigment gene or regulator. NOT randomly chosen markers genome-wide.
Candidate gene approach

Red progeny have RFLP pattern like red parent

Unpigmented progeny have RFLP pattern like unpigmented parent

But if you can beat multiple testing, why not do the whole genome...
Testing for linkage doesn’t always mean counting recombinants.

Back to week 4

Qualitative but polygenic

Two loci.
Need one dominant allele at each locus to get phenotype.

A simulated cross: test one locus

Flower color

Genotype at marker close to A locus

Need one dominant allele at each locus to get phenotype.
A simulated cross: test one locus

Flower color

Genotype at marker close to A locus

Purple flowers result from AA or Aa.

No need to count recombinants
No need to count recombinants

Flower color

χ² = Σ(O - E)² / E

Genotype at marker close to A locus

“A weak locus”

Because A locus by itself is not the whole story, studying it in isolation gives only weak statistical significance.

χ² = Σ(O - E)² / E

Many traits—cancers, cleft palate, high blood pressure—fit this description.
Multiple loci underlie many yes-or-no traits

“Threshold model”

Affected sib pair method

Affected sib pair method

What is probability of this by chance?

A. 1/4
B. 1/2
C. 1/8
D. 1/3
What is probability of both kids getting 2 from mom or both kids getting 3 from mom?

A. 1/4
B. 1/2
C. 1/8
D. 1/3

\((\frac{1}{2})^2 \times (\frac{1}{2})^2 + (\frac{1}{2})^2 \times (\frac{1}{2})^2\)

\(\text{Prob of both getting 2} + \text{Prob of both getting 3}\)

Expected under null

\(\chi^2 = \sum \frac{(O - E)^2}{E}\)

Test for significant allele sharing.
**Affected sib pair method**

<table>
<thead>
<tr>
<th>Sib pairs</th>
<th>Observed</th>
<th>Expected under null</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same allele</td>
<td>2</td>
<td>(1/2)*2</td>
</tr>
<tr>
<td>Different allele</td>
<td>0</td>
<td>(1/2)*2</td>
</tr>
</tbody>
</table>

Doesn’t require you to know dominant or recessive, one locus or two, ...

**Quantitative traits**

Unlike cystic fibrosis and Huntington’s disease, most traits are not yes-or-no.
Unlike cystic fibrosis and Huntington's disease, most traits are not yes-or-no.

E.g. blood pressure.
Environment and error

What if…

Salt water
Plain water

• Time of day
• Change in cage-mates
• Age
• Reproductive cycle

What if…

Exact same mouse, every day for 6 mo
What if…

Many clones/identical twins

What if…

“Experimental error”
+ random variation

Many clones/identical twins

- Time of day
- Change in cage-mates
- Age
- Reproductive cycle