ESP
Your friend claims he can predict the future
You flip a coin 5 times, and he’s right on 4
Is your friend psychic?

Two Hypotheses

Hypothesis
A theory about how the world works
Proposed as an explanation for data
Posed as statement about population parameters

Psychic
Some ability to predict future
Not perfect, but better than chance

Luck
Random chance
Right half the time, wrong half the time

Hypothesis testing
A method that uses inferential statistics to decide which of two hypotheses the data support

Likelihood

Likelihood
Probability distribution of a statistic, according to each hypothesis
If result is likely according to a hypothesis, we say data “support” or “are consistent with” the hypothesis

Likelihood for $f(correct)$
Psychic: hard to say; how psychic?
Luck: can work out exactly; 50/50 chance each time

Likelihood According to Luck

<table>
<thead>
<tr>
<th>Y</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1/32</td>
</tr>
<tr>
<td>4</td>
<td>5/32</td>
</tr>
<tr>
<td>3</td>
<td>10/32</td>
</tr>
<tr>
<td>2</td>
<td>10/32</td>
</tr>
<tr>
<td>1</td>
<td>5/32</td>
</tr>
<tr>
<td>0</td>
<td>1/32</td>
</tr>
</tbody>
</table>
Binomial Distribution

Binary data
A set of two-choice outcomes, e.g. yes/no, right/wrong

Binomial variable
A statistic for binary samples
Frequency of “yes” / “right” / etc.

Binomial distribution
Probability distribution for a binomial variable
Gives probability for each possible value, from 0 to \( n \)

A family of distributions
Like Normal (need to specify mean and SD)
\( n \): number of observations (sample size)
\( q \): probability correct each time

Formula (optional):
\[
p(f) = q^f (1-q)^{n-f} \frac{n!}{f!(n-f)!}
\]

Binomial Test
Hypothesis testing for binomial statistics

Null hypothesis
Some fixed value for \( q \), usually \( q = .5 \)
Nothing interesting going on; blind chance (no ESP)

Alternative hypothesis
\( q \) equals something else
One outcome more likely than expected by chance (ESP)

Goal: Decide which hypothesis the data support

Strategy
Find likelihood distribution for \( f(Y) \) according to null hypothesis
Compare actual result to this distribution
If actual result is too extreme, reject null hypothesis and accept alternative hypothesis

“Innocent until proven guilty”
Believe null hypothesis unless compelling evidence to rule it out
Only accept ESP if luck can’t explain the data

Testing for ESP
Null hypothesis: Luck, \( q = .5 \)
Alternative hypothesis: ESP, \( q > .5 \)

Need to decide rules in advance
If too extreme, abandon luck and accept ESP

How unlikely before we will give up on Luck?

Where to draw cutoff?

Critical value: value our statistic must exceed to reject null hypothesis (Luck)

Another Example: Treatment Evaluation
Do people tend to get better with some treatment?
Less depression, higher WBC, better memory, etc.
Measure who improves and who worsens
Want more people better off than worse
Sign test
Ignore magnitude of change; just direction
Same logic as other binomial tests
Count number of patients improved
Compare to probabilities according to chance

Structure of a Binomial Test

<table>
<thead>
<tr>
<th>Binary data</th>
<th>Each coin prediction is Correct or Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each patient is Better or Worse</td>
<td>Probability each patient will improve</td>
</tr>
<tr>
<td>Population parameter $q$</td>
<td>Probability each guess will be correct</td>
</tr>
<tr>
<td>Probability each patient will improve</td>
<td>No effect of treatment</td>
</tr>
<tr>
<td>Null hypothesis, usually $q = .5$</td>
<td>No ESP</td>
</tr>
<tr>
<td>Better or worse equally likely</td>
<td>Correct and incorrect equally likely</td>
</tr>
<tr>
<td>Alternative hypothesis, here $q &gt; .5$</td>
<td>Effective treatment; more people improve</td>
</tr>
<tr>
<td>Work out likelihood according to null hypothesis</td>
<td>ESP; guess right more often than chance</td>
</tr>
<tr>
<td>Probability distribution for $f$(improve)</td>
<td>Probability distribution for $f$(correct)</td>
</tr>
<tr>
<td>Compare actual result to these probabilities</td>
<td>If more improve than likely by chance,</td>
</tr>
<tr>
<td></td>
<td>accept treatment is useful</td>
</tr>
<tr>
<td>Need to decide critical value</td>
<td>If more correct than likely by chance,</td>
</tr>
<tr>
<td></td>
<td>abandon luck and accept ESP</td>
</tr>
<tr>
<td>How many patients must improve?</td>
<td>How many times correct?</td>
</tr>
</tbody>
</table>

Errors
Whatever the critical value, there will be errors
All values 0 to $n$ are possible under null hypothesis
Even 20/20 happens once in 1,048,576 times
Can only minimize how often errors occur
Two kinds of errors:
Type I error
Null hypothesis is true, but we reject it
Conclude a useless treatment is effective
Type II error
Null hypothesis is false, but we don’t reject it
Don’t recognize when a treatment is effective

Critical Value and Error Rates
Increasing critical value reduces Type I error rate but increases Type II error rate (and vice versa)
So, how do we decide critical value?
Two principles
Type I errors are more important to avoid
Can’t figure out Type II error rate anyway
Strategy
Decide how many Type I errors are acceptable
Choose critical value accordingly
Controlling Type I Error

Type I error rate
- Proportion of times, when null hypothesis is true, that we mistakenly reject it
- Fraction of bogus treatments that we conclude are effective
- Type I error rate equals total probability beyond the critical value, according to null hypothesis

Strategy
- Decide what Type I error rate we want to allow
- Pick critical value accordingly

Alpha level ($\alpha$)
- Chosen Type I error rate
- Usually .05 in Psychology
- Determines critical value

Summary of Hypothesis Testing
- Determine Null and Alternative Hypotheses
  - Competing possibilities about a population parameter
  - Null is always precise; usually means “no effect”
- Find probability distribution of test statistic according to null hypothesis
  - Likelihood of the statistic under that hypothesis
- Choose acceptable rate of Type I errors ($\alpha$)
- Pick critical value of test statistic based on $\alpha$
  - Under Null, probability of a result past critical value equals $\alpha$
- Compare actual result to critical value
  - If more extreme, reject Null as unable to explain data
  - Otherwise, stick with Null because it's an adequate explanation