Interpreting Evidence

Elijah J. Spina
What does this mean?
Important Questions

• What is scientific evidence?

• How is evidence generated?

• How can evidence be used to test hypotheses?
What is science?

The systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles based on these facts. The organized body of knowledge that is derived from such observations and that can be verified or tested by further investigation.

Academic Press Dictionary of Science & Technology

Science consists simply of the formulation and testing of hypotheses based on observational evidence; experiments are important where applicable, but their function is merely to simplify observation by imposing controlled conditions.

Robert H. Dott, Jr., and Henry L. Batten, *Evolution of the Earth*

Science alone of all the subjects contains within itself the lesson of the danger of belief in the infallibility of the greatest teachers in the preceding generation . . . As a matter of fact, I can also define science another way: Science is the belief in the ignorance of experts.

Richard Feynman
The Scientific Method

1. **Observation**
   - Conduct experiment

2. **Hypothesis**
   - Make predictions using hypothesis
   - Revise or confirm hypothesis

3. **Prediction**
   - Devise experiment to test predictions
Terms Used to Describe the “Nature of Science”

**Fact**: An observation that has been repeatedly confirmed and for all practical purposes is accepted as "true." Truth in science, however, is never final, and what is accepted as a fact today may be modified or even discarded tomorrow.

**Hypothesis**: A falsifiable statement about the natural world.

**Law**: A descriptive generalization about how some aspect of the natural world behaves under stated circumstances.

**Theory**: A well-substantiated explanation of some aspect of the natural world that can incorporate facts, laws, inferences, and tested hypotheses.

Modified from National Academy of Sciences, 1999
Scientific Theories

The contention that evolution should be taught as a "theory, not as a fact" confuses the common use of these words with the scientific use. In science, theories do not turn into facts through the accumulation of evidence. Rather, theories are the end points of science. They are understandings that develop from extensive observation, experimentation, and creative reflection. They incorporate a large body of scientific facts, laws, tested hypotheses, and logical inferences. In this sense, evolution is one of the strongest and most useful scientific theories we have.
Scientific Theories

General Theory of Relativity

Cell Theory
Paradigms

Thomas Kuhn (1922-1996)
*The Structure of Scientific Revolutions*

“Universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of practitioners.”

Wikipedia: The set of practices that define a scientific discipline at any particular period of time.
Types of Reasoning

Inductive logic: “bottom up”
• Inferring a general conclusion from individual observations

Deductive logic: “top down”
• Narrowing down general cases to a specific conclusion
Evidence of Causality

Necessity:
• If \( x \) is a necessary cause of \( y \),
• Then the presence of \( y \) necessarily implies the presence of \( x \).
• The presence of \( x \), however, does not imply that \( y \) will occur.

Sufficiency:
• If \( x \) is a sufficient cause of \( y \),
• Then the presence of \( x \) necessarily implies the presence of \( y \).
• However, another cause \( z \) may alternatively cause \( y \).
• Thus the presence of \( y \) does not imply the presence of \( x \).
Forms of Argument

Modus ponens:
- If today is Tuesday, then John will go to work.
- Today is Tuesday.
- Therefore, John will go to work.

Modus tollens:
- If the watch-dog detects an intruder, the watch-dog will bark.
- The watch-dog did not bark
- Therefore, no intruder was detected by the watch-dog.
Reductionism

The whole is just the sum of the parts.

“The universe can be understood by abstracting and breaking down each system into component parts that are ever more basic and fundamental. Reductionism exposes the nature of a system whereas scientific hypothesis testing allows for the synthesis of scientific truth.”

Is this always the case?

- Larger scale organization can influence smaller scales.
- Feedback loops create structure at a given level, independently of details at a lower level of organization.
Controls Make The World Go Round

Controls are “anti-experiments”:

- They represent a null hypothesis.
- The answer should already be known.
- Nothing is being tested. It’s either going to ‘work’ (positive control) or it’s going to have no effect (negative control).
Activity
Discuss the examples on the handout
Think about types of reasoning and causality
Preferably work in pairs or groups of 3
Statistical Inference

The process of drawing conclusions from data that are subject to random variation.

• Central to inductive arguments.

• Paradigms:
  Frequentist: sharp decision rules for propositions
  Bayesian: degrees of belief in a proposition
**DID THE SUN JUST EXPLODE?**
(It's night, so we're not sure)

This neutrino detector measures whether the sun has gone nova.

Then, it rolls two dice. If they both come up six, it lies to us.
Otherwise, it tells the truth.

Let's try.

**Detecter! Has the sun gone nova?**

Roll:

**YES.**

**Frequentist Statistician:**

The probability of this result happening by chance is \( \frac{1}{36} = 0.027 \). Since \( p < 0.05 \), I conclude that the sun has exploded.

**Bayesian Statistician:**

BET YOU $50 IT HASN'T.
**Statistical vs. Practical Significance**

**PROBABLE CAUSE**

A P value measures whether an observed result can be attributed to chance. But it cannot answer a researcher’s real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

**Before the experiment**

The plausibility of the hypothesis — the odds of it being true — can be estimated from previous experiments, conjectured mechanisms and other expert knowledge. Three examples are shown here.

**The measured P value**

A value of 0.05 is conventionally deemed ‘statistically significant’; a value of 0.01 is considered ‘very significant’.

**After the experiment**

A small P value can make a hypothesis more plausible, but the difference may not be dramatic.

---

Illustrating A “Long Shot”

JELLY BEANS
CAUSE ACNE!

SCIENTISTS!
INVESTIGATE!

BUT WE'RE PLAYING MINECRAFT!
...FINE.

WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).

THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.

SCIENTISTS!

BUT MINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE ($P < 0.05$).

WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE ($P > 0.05$).
Green jelly beans linked to acne!

95% confidence

Only 5% chance of coincidence!

Scientists...
## Common Types of Effect Size

<table>
<thead>
<tr>
<th>Index</th>
<th>Description</th>
<th>Effect Size</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between groups</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Cohen’s $d^a$          | $d = M_1 - M_2 / s$  
$M_1 - M_2$ is the difference between the group means ($M$); $s$ is the standard deviation of either group | Small 0.2  
Medium 0.5  
Large 0.8  
Very large 1.3 | Can be used at planning stage to find the sample size required for sufficient power for your study |
| Odds ratio (OR)        | Group 1 odds of outcome  
Group 2 odds of outcome  
If OR = 1, the odds of outcome are equally likely in both groups | Small 1.5  
Medium 2  
Large 3 | For binary outcome variables  
Compares odds of outcome occurring from one intervention vs another |
| Relative risk or risk ratio (RR) | Ratio of probability of outcome in group 1 vs group 2;  
If RR = 1, the outcome is equally probable in both groups | Small 2  
Medium 3  
Large 4 | Compares probabilities of outcome occurring from one intervention to another |
| **Measures of association** |                                                  |                     |                                                                          |
| Pearson’s $r$ correlation | Range, -1 to 1  
| | Small $\pm$0.2  
Medium $\pm$0.5  
Large $\pm$0.8 | Measures the degree of linear relationship between two quantitative variables |
| $r^2$ coefficient of determination | Range, 0 to 1; Usually expressed as percent | Small 0.04  
Medium 0.25  
Large 0.64 | Proportion of variance in one variable explained by the other |
## Error Bars

<table>
<thead>
<tr>
<th>Error bar</th>
<th>Type</th>
<th>Description</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>Descriptive</td>
<td>Amount of spread between the extremes of the data</td>
<td>Highest data point minus the lowest</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>Descriptive</td>
<td>Typical or (roughly speaking) average difference between the data points and their mean</td>
<td>$SD = \sqrt{\frac{\sum (X - M)^2}{n - 1}}$</td>
</tr>
<tr>
<td>Standard error (SE)</td>
<td>Inferential</td>
<td>A measure of how variable the mean will be, if you repeat the whole study many times</td>
<td>$SE = \frac{SD}{\sqrt{n}}$</td>
</tr>
<tr>
<td>Confidence interval (CI), usually 95% CI</td>
<td>Inferential</td>
<td>A range of values you can be 95% confident contains the true mean</td>
<td>$M \pm t_{(n-1)} \times SE$, where $t_{(n-1)}$ is a critical value of $t$. If $n$ is 10 or more, the 95% CI is approximately $M \pm 2 \times SE$.</td>
</tr>
</tbody>
</table>
The Overlap Rule
Try out the overlap rule

Means w/ SE bars
n = 3

Can you estimate $p$-values?
- Between groups?
- Within groups?
Important Questions

• What is scientific evidence?
  • There are multiple levels, it depends on the questions and context of interpretation.

• How is evidence generated?
  • Understanding the “nature of science”.
  • Applying the scientific method.

• How can hypotheses be tested?
  • Building support (induction)
  • Ruling things out (deduction)
  • Statistical inference
Download Locations

ejsbio.weebly.com

labs.mcdb.ucsb.edu/smith/william/members/spina