Statistical Methods and Data Analysis I
Lecture 3: Hypothesis Testing Fundamentals

Oleg Goldshmidt
oleg.goldshmidt@post.idc.ac.il

Arison School of Business
Interdisciplinary Center (IDC)
Herzliya, Israel

March 10, 2019

Variation

Variation

- Everything varies:
  - with time
  - in space
  - from one object to another
  - when some parameters change

Variation

- Everything varies:
  - with time
  - in space
  - from one object to another
  - when some parameters change
- Background variation vs. *interesting variation*:
  - this is why you need statistics — to distinguish between the two!

Variation

- Everything varies:
  - with time
  - in space
  - from one object to another
  - when some parameters change
- Background variation vs. *interesting variation*:
  - this is why you need statistics — to distinguish between the two!
- Key concept: variation that one can expect to occur *by chance*, in the background —
  - if we observe more variation we say that the effect is *"statistically significant"*
  - if we do not observe more variation then the result is *"not statistically significant"*

Variation

- Everything varies:
  - with time
  - in space
  - from one object to another
  - when some parameters change
- Background variation vs. *interesting variation*:
  - this is why you need statistics — to distinguish between the two!
- Key concept: variation that one can expect to occur *by chance*, in the background —
  - if we observe more variation we say that the effect is *"statistically significant"*
  - if we do not observe more variation then the result is *"not statistically significant"*
- Warning: "not statistically significant" does **not** mean "not important"
  - it means there is no evidence that it is, within the limitations of a particular experiment or data set
  - absence of evidence is not evidence of absence!

IDC HERZLIYA

Variation

- Everything varies:
  - with time
  - in space
  - from one object to another
  - when some parameters change
- Background variation vs. *interesting variation*:
  - this is why you need statistics — to distinguish between the two!
- Key concept: variation that one can expect to occur *by chance*, in the background —
  - if we observe more variation we say that the effect is *"statistically significant"*
  - if we do not observe more variation then the result is *"not statistically significant"*
- Warning: "not statistically significant" does **not** mean "not important"
  - it means there is no evidence that it is, within the limitations of a particular experiment or data set
  - absence of evidence is not evidence of absence!
- Negative results are as important as positive statements!

Statistical Significance

Statistically significant result

. . . is *unlikely* to occur by chance under a specific set of assumptions called “null hypothesis”.

Statistical Significance

Statistically significant result

. . . is *unlikely* to occur by chance under a specific set of assumptions called “null hypothesis”.

We need to define “unlikely”...

Statistical Significance

Statistically significant result

. . . is *unlikely* to occur by chance under a specific set of assumptions called "null hypothesis".

We need to define "unlikely"...

Null hypothesis (H_0)

A general statement (default position) that there is no relationship between two measured phenomena, or no association between the observed parameters (*"nothing is happening"*).

Statistical Significance

Statistically significant result

. . . is *unlikely* to occur by chance under a specific set of assumptions called "null hypothesis".

We need to define "unlikely"...

Null hypothesis (H_0)

A general statement (default position) that there is no relationship between two measured phenomena, or no association between the observed parameters (*"nothing is happening"*).

Statistical inference

- If the observed data are *significantly unlikely* to have occurred if H_0 were true then H_0 is **rejected** and an "alternative hypothesis" (H_A — *"something is happening"*) is accepted.
- If the observed data are consistent with H_0 then H_0 is not rejected.
- In neither case can we say that H_0 or H_A is **proven** (cf. *presumption of innocence*).

Creating Hypotheses

What constitutes a good hypothesis?

A good hypothesis is falsifiable. (Karl Popper)

Creating Hypotheses

What constitutes a good hypothesis?

A good hypothesis is falsifiable. (Karl Popper)

This is an essential foundation of the "scientific method" and of all **empirical sciences!**

Creating Hypotheses

What constitutes a good hypothesis?

A good hypothesis is falsifiable. (Karl Popper)

This is an essential foundation of the "scientific method" and of all **empirical sciences!**

Hempel's Paradox (Carl Hempel)

It takes just *one example* to the contrary to **reject** a hypothesis, but not even *an infinite number* of supporting examples can prove a hypothesis.

Creating Hypotheses

What constitutes a good hypothesis?

A good hypothesis is falsifiable. (Karl Popper)

This is an essential foundation of the "scientific method" and of all **empirical sciences!**

Hempel's Paradox (Carl Hempel)

It takes just *one example* to the contrary to **reject** a hypothesis, but not even *an infinite number* of supporting examples can prove a hypothesis.

Hypothesis: **all ravens are black**

Logically equivalent statement (double negation): if something is **not** black then it is not a raven

Creating Hypotheses

What constitutes a good hypothesis?

A good hypothesis is falsifiable. (Karl Popper)

This is an essential foundation of the "scientific method" and of all **empirical sciences!**

Hempel's Paradox (Carl Hempel)

It takes just *one example* to the contrary to **reject** a hypothesis, but not even *an infinite number* of supporting examples can prove a hypothesis.

Hypothesis: **all ravens are black**

Logically equivalent statement (double negation): if something is **not** black then it is not a raven

Supporting evidence:

Credit: Wikipedia

The Prosecutor's Fallacy

We still need to define "likely", "unlikely", and "beyond reasonable doubt" (recalling that the "legal null hypothesis" is presumption of innocence).

The Prosecutor's Fallacy

We still need to define "likely", "unlikely", and "beyond reasonable doubt" (recalling that the "legal null hypothesis" is presumption of innocence).

Credit: Everett Collection/Shutterstock

Legal case: a murder trial

Forensics found DNA matching the accused on the murder weapon.
Prosecution: This is damning evidence, beyond reasonable doubt.
The jury gasp... Then nod thoughtfully...

The Prosecutor's Fallacy

We still need to define "likely", "unlikely", and "beyond reasonable doubt" (recalling that the "legal null hypothesis" is presumption of innocence).

Credit: Everett Collection/Shutterstock

Legal case: a murder trial

Forensics found DNA matching the accused on the murder weapon.
Prosecution: This is damning evidence, beyond reasonable doubt.
The jury gasp... Then nod thoughtfully...

Defence: Surely it may be just a coincidence!

The Prosecutor's Fallacy

We still need to define "likely", "unlikely", and "beyond reasonable doubt" (recalling that the "legal null hypothesis" is presumption of innocence).

Credit: Everett Collection/Shutterstock

Legal case: a murder trial

Forensics found DNA matching the accused on the murder weapon.
Prosecution: This is damning evidence, beyond reasonable doubt.
The jury gasp... Then nod thoughtfully...

Defence: Surely it may be just a coincidence!
Prosecution: *Science* tells us the chances of such coincidence are minuscule!

The Prosecutor's Fallacy

We still need to define "likely", "unlikely", and "beyond reasonable doubt" (recalling that the "legal null hypothesis" is presumption of innocence).

Credit: Everett Collection/Shutterstock

Legal case: a murder trial

Forensics found DNA matching the accused on the murder weapon.
Prosecution: This is damning evidence, beyond reasonable doubt.
The jury gasp... Then nod thoughtfully...

Defence: Surely it may be just a coincidence!
Prosecution: *Science* tells us the chances of such coincidence are minuscule!
Defence: Objection! The Prosecution's statement is **not** scientific! *Science* would be quantitative...

The Prosecutor's Fallacy

We still need to define "likely", "unlikely", and "beyond reasonable doubt" (recalling that the "legal null hypothesis" is presumption of innocence).

Credit: Everett Collection/Shutterstock

Legal case: a murder trial

Forensics found DNA matching the accused on the murder weapon.
Prosecution: This is damning evidence, beyond reasonable doubt.
The jury gasp... Then nod thoughtfully...

Defence: Surely it may be just a coincidence!
Prosecution: *Science* tells us the chances of such coincidence are minuscule!
Defence: Objection! The Prosecution's statement is **not** scientific! *Science* would be quantitative...
Prosecution: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that the probability of a chance DNA match is **ONE IN A MILLION!**

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**
Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**

Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???

Defence: The *scientific fact* that the probability of a chance DNA match is one in a million is not in doubt at all. But Defense does indeed intend to prove that it has no relevance to the case!

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**

Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???

Defence: The *scientific fact* that the probability of a chance DNA match is one in a million is not in doubt at all. But Defense does indeed intend to prove that it has no relevance to the case!

Prosecution: **???** *Jury:* **<GASP!!!>**

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**
Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???
Defence: The *scientific fact* that the probability of a chance DNA match is one in a million is not in doubt at all. But Defense does indeed intend to prove that it has no relevance to the case!
Prosecution: **???** *Jury:* **<GASP!!!>**

Defence: Let's draw a little table...

Hypothesis	Match	No Match
Guilty		
Innocent		

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**
Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???
Defence: The *scientific fact* that the probability of a chance DNA match is one in a million is not in doubt at all. But Defense does indeed intend to prove that it has no relevance to the case!
Prosecution: **???** *Jury:* **<GASP!!!>**

Defence: Let's draw a little table...
Defence: There is only one murderer, right? We believe in our DNA science: either there is a match, or there is none.

Hypothesis	Match	No Match
Guilty	1	0
Innocent		

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**
Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???
Defence: The *scientific fact* that the probability of a chance DNA match is one in a million is not in doubt at all. But Defense does indeed intend to prove that it has no relevance to the case!
Prosecution: **???** *Jury:* **<GASP!!!>**

Defence: Let's draw a little table...
Defence: There is only one murderer, right? We believe in our DNA science: either there is a match, or there is none.
Defence: Now, there are $\sim 4,000,000$ people in the Greater Tel Aviv / Gush Dan area...

Hypothesis	Match	No Match
Guilty	1	0
Innocent		4,000,000

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**
Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???
Defence: The *scientific fact* that the probability of a chance DNA match is one in a million is not in doubt at all. But Defense does indeed intend to prove that it has no relevance to the case!
Prosecution: **???** *Jury:* **<GASP!!!>**

Defence: Let's draw a little table...
Defence: There is only one murderer, right? We believe in our DNA science: either there is a match, or there is none.
Defence: Now, there are $\sim 4,000,000$ people in the Greater Tel Aviv / Gush Dan area... With the probability of chance DNA match as Prosecution stated...

Hypothesis	Match	No Match
Guilty	1	0
Innocent		4,000,000

$\overleftarrow{p = 1/1,000,000}$

The Prosecutor's Fallacy (Cont.)

Credit: https://imdb.com

Defence: Your Honor, Ladies and Gentlemen of the Jury, it is a *scientific fact* that prosecution is talking **UTTER NONSENSE!**
Prosecution: Objection! Does the Defence Counsel intend to dispute a well-known *scientific fact*???
Defence: The *scientific fact* that the probability of a chance DNA match is one in a million is not in doubt at all. But Defense does indeed intend to prove that it has no relevance to the case!
Prosecution: **???** *Jury:* **<GASP!!!>**

Defence: Let's draw a little table...
Defence: There is only one murderer, right? We believe in our DNA science: either there is a match, or there is none.
Defence: Now, there are $\sim 4,000,000$ people in the Greater Tel Aviv / Gush Dan area... With the probability of chance DNA match as Prosecution stated... **There is only a 20% chance that my client is guilty!**

Hypothesis	Match	No Match
Guilty	1	0
Innocent	4	4,000,000

$\overleftarrow{p = 1/1,000,000}$

The Prosecutor's Fallacy: Conditional Probability

Conditional Probability

Probability of event A occurring when we *know* that event B has occurred. Notation: $P(A|B)$.

Hypothesis	Match	No Match
Guilty	1	0
Innocent	4	4,000,000

$\overleftarrow{p = 1/1,000,000}$

The Prosecutor's Fallacy: Conditional Probability

Conditional Probability

Probability of event A occurring when we *know* that event B has occurred. Notation: $P(A|B)$.

Hypothesis	Match	No Match
Guilty	1	0
Innocent	4	4,000,000

$\overleftarrow{p = 1/1,000,000}$

Prosecutor's Statement:

The conditional probability of a DNA match *given that the accused is innocent* is **1:1,000,000!**

The Prosecutor's Fallacy: Conditional Probability

Conditional Probability

Probability of event A occurring when we *know* that event B has occurred. Notation: $P(A|B)$.

Hypothesis	Match	No Match
Guilty	1	0
Innocent	4	4,000,000

$\overleftarrow{p = 1/1,000,000}$

Prosecutor's Statement:

The conditional probability of a DNA match *given that the accused is innocent* is **1:1,000,000!**

This is a statement about the efficacy of DNA testing, **not** *of guilt or innocence of the accused!*

The Prosecutor's Fallacy: Conditional Probability

Conditional Probability

Probability of event A occurring when we *know* that event B has occurred. Notation: $P(A|B)$.

Hypothesis	Match	No Match
Guilty	1	0
Innocent	4	4,000,000

$\overleftarrow{p = 1/1,000,000}$

Prosecutor's Statement:

The conditional probability of a DNA match *given that the accused is innocent* is **1:1,000,000!**

This is a statement about the efficacy of DNA testing, **not** *of guilt or innocence of the accused!*

What Prosecution should have stated:

The conditional probability of the accused being **guilty** *given that there is a DNA match* is . . . **20%!**

The Prosecutor's Fallacy: Conditional Probability

Conditional Probability

Probability of event A occurring when we *know* that event B has occurred. Notation: $P(A|B)$.

Hypothesis	Match	No Match
Guilty	1	0
Innocent	4	4,000,000

$\overleftarrow{p = 1/1,000,000}$

Formally:

$$P(Match|Innocent) = 1/1000000$$
$$P(Guilt|Match) = 0.2$$

Prosecutor's Statement:

The conditional probability of a DNA match *given that the accused is innocent* is **1:1,000,000!**

This is a statement about the efficacy of DNA testing, **not** *of guilt or innocence of the accused!*

What Prosecution should have stated:

The conditional probability of the accused being **guilty** *given that there is a DNA match* is . . . **20%!**

Hypothesis Testing: p-value

Testing H_0	(match) Reject H_0	(no match) Do not reject H_0
H_0 false (guilty)	True positive (correct decision)	False negative (Type II error)
H_0 true (innocent)	False positive (Type I error)	True negative (correct decision)

p-value

The probability of **rejecting** H_0 *by mistake*.

In our courtroom example: p-value = 1/1,000,000.

Hypothesis Testing: Is The Hypothesis True?

We do **not** *want to know the p-value! We want to know whether our hypothesis is true or false!*

Hypothesis Testing: Is The Hypothesis True?

We do **not** *want to know the p-value! We want to know whether our hypothesis is true or false!*

If we reject H_0, what is the chance that H_0 *is actually false* (and H_A is true)?

Hypothesis Testing: Is The Hypothesis True?

We do **not** *want to know the p-value! We want to know whether our hypothesis is true or false!*

If we reject H_0, what is the chance that H_0 *is actually false* (and H_A is true)?

Assume we performed 1000 experiments:

Hypothesis	Reject H_0	Do not reject H_0
H_0 false	**?**	**?**
H_0 true	50	950

Hypothesis Testing: Is The Hypothesis True?

We do **not** *want to know the p-value! We want to know whether our hypothesis is true or false!*

If we reject H_0, what is the chance that H_0 *is actually false* (and H_A is true)?

Assume we performed 1000 experiments:

Hypothesis	Reject H_0	Do not reject H_0
H_0 false	?	?
H_0 true	50	950

Our p-value is $p = 50/1000 = 5\%$.

Hypothesis Testing: Is The Hypothesis True?

We do **not** *want to know the p-value! We want to know whether our hypothesis is true or false!*

If we reject H_0, what is the chance that H_0 *is actually false* (and H_A is true)?

Assume we performed 1000 experiments:

Hypothesis	Reject H_0	Do not reject H_0
H_0 false	**?**	**?**
H_0 true	50	950

Our p-value is $p = 50/1000 = 5\%$.

- How likely is it that our H_A is true *before the experiment*?
  - *a priori* probability $P(H_A)$

Hypothesis Testing: Is The Hypothesis True?

We do **not** *want to know the p-value! We want to know whether our hypothesis is true or false!*

If we reject H_0, what is the chance that H_0 *is actually false* (and H_A is true)?

Assume we performed 1000 experiments:

Hypothesis	Reject H_0	Do not reject H_0
H_0 false	2	998
H_0 true	50	950

Our p-value is $p = 50/1000 = 5\%$.

- How likely is it that our H_A is true *before the experiment*?
  - *a priori* probability $P(H_A)$
- A really crazy hypothesis — $P(H_A) = 1 : 500$:
  - *a posteriori* probability $P(H_A|reject) = 2/52 \approx 3.85\%$

Hypothesis Testing: Is The Hypothesis True?

We do **not** *want to know the p-value! We want to know whether our hypothesis is true or false!*

If we reject H_0, what is the chance that H_0 *is actually false* (and H_A is true)?

Assume we performed 1000 experiments:

Hypothesis	Reject H_0	Do not reject H_0
H_0 false	500	500
H_0 true	50	950

Our p-value is $p = 50/1000 = 5\%$.

- How likely is it that our H_A is true *before the experiment*?
  - *a priori* probability $P(H_A)$
- A really crazy hypothesis — $P(H_A) = 1 : 500$:
  - *a posteriori* probability $P(H_A|reject) = 2/52 \approx 3.85\%$
- A plausible (fifty-fifty) hypothesis — $P(H_A) = 1 : 1$:
  - *a posteriori* probability $P(H_A|reject) = 500/550 \approx 90.91\%$

The Meaning of p-value

What p-value is **not**:

The p-value is **not** how likely you think your alternative hypothesis is to be **true**...

The Meaning of p-value

What p-value is **not**:

The p-value is **not** how likely you think your alternative hypothesis is to be **true**...

What p-value **is**:

It is how likely you think the result would have been if your alternative hypothesis were **false**!

The Meaning of p-value

What p-value is **not**:

The p-value is **not** how likely you think your alternative hypothesis is to be **true**...

What p-value **is**:

It is how likely you think the result would have been if your alternative hypothesis were **false**!

Quantifying "statistical significance":

The smaller the p-value the more "statistically significant" your result is.
If your result is very unlikely under H_0 (i.e., if H_A is false) then "something is happening" ("there is an observable effect", etc.) *with high statistical significance*.
NB: This still does not tell you how likely it is that H_A is true...

The Meaning of p-value

What p-value is **not**:

The p-value is **not** how likely you think your alternative hypothesis is to be **true**...

What p-value **is**:

It is how likely you think the result would have been if your alternative hypothesis were **false**!

Quantifying "statistical significance":

The smaller the p-value the more "statistically significant" your result is.
If your result is very unlikely under H_0 (i.e., if H_A is false) then "something is happening" ("there is an observable effect", etc.) *with high statistical significance*.
NB: This still does not tell you how likely it is that H_A is true...

How low should p-value be for a "significant result"?

Have patience...

More Examples

- In the courtroom:
  - H_0: the defendant is innocent; H_A: the defendant is guilty...
  - A p-value of 1/1000000 does not mean that the chances are 1000000 to 1 that the defendant is guilty (H_A is true)...
  - It means that if the defendant is innocent (H_A is false) there is a one in a million chance that we will get a DNA match on the crime scene...
  - The jury does not assert innocence — the verdict is “guilty” or “not guilty”...
    - “there is not enough evidence that the defendant is guilty” vs. “the defendant is innocent”
    - hopefully the investigation checks for the latter...

More Examples

- In the courtroom:
  - H_0: the defendant is innocent; H_A: the defendant is guilty...
  - A p-value of 1/1000000 does not mean that the chances are 1000000 to 1 that the defendant is guilty (H_A is true)...
  - It means that if the defendant is innocent (H_A is false) there is a one in a million chance that we will get a DNA match on the crime scene...
  - The jury does not assert innocence — the verdict is "guilty" or "not guilty"...
    - "there is not enough evidence that the defendant is guilty" vs. "the defendant is innocent"
    - hopefully the investigation checks for the latter...
- In science:
  - H_0: smoking does not increase cancer risk; H_A: it does... *Assume empirical p-value is 0.005.*
  - *Does this mean that with empirical probability 99.5% smoking increases cancer risk?*
  - *Yes, if you smoke your chances to get cancer increase by a factor of 200!*
  - *If smoking in unrelated to cancer there is just 1 chance in 200 that we would get our experimental results at random.*

More Examples

- In the courtroom:
  - H_0: the defendant is innocent; H_A: the defendant is guilty...
  - A p-value of 1/1000000 does not mean that the chances are 1000000 to 1 that the defendant is guilty (H_A is true)...
  - It means that if the defendant is innocent (H_A is false) there is a one in a million chance that we will get a DNA match on the crime scene...
  - The jury does not assert innocence — the verdict is "guilty" or "not guilty"...
    - "there is not enough evidence that the defendant is guilty" vs. "the defendant is innocent"
    - hopefully the investigation checks for the latter...
- In science:
  - H_0: smoking does not increase cancer risk; H_A: it does... *Assume empirical p-value is 0.005.*
  - *Does this mean that with empirical probability 99.5% smoking increases cancer risk?*
  - *Yes, if you smoke your chances to get cancer increase by a factor of 200!*
  - **If smoking in unrelated to cancer there is just 1 chance in 200 that we would get our experimental results at random.**