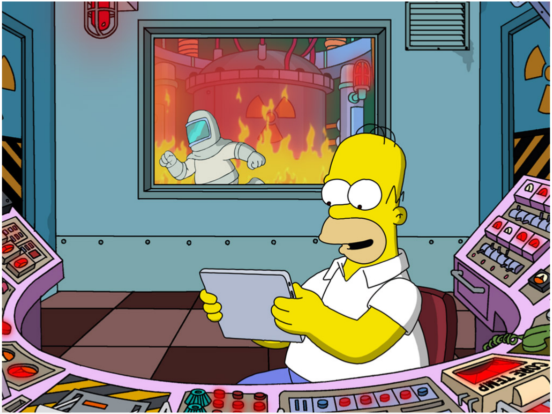
Statistical Hypothesis Testing

Tuesday 29th November, 2016

Motivation - Control Room

- new job
- night watch monitor at a nuclear power plant
- task: raise the alarm if anything is out of the ordinary

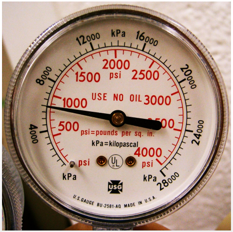
Motivation - Control Room

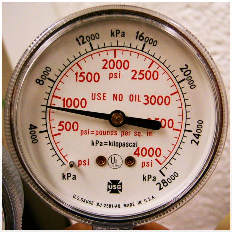


Motivation - Control Room

Israeltech challenge

- new job
- night watch monitor at a nuclear power plant
- task: raise the alarm if anything is out of the ordinary
- Specifically - watch two meters - heat and pressure





Motivation - Control Room

Israeltech challenge

- new job
- night watch monitor at a nuclear power plant
- task: raise the alarm if anything is out of the ordinary
- Specifically - watch two meters - heat and pressure
- on your first night - various readings
- when and if do you raise the alarm?

discuss

Motivation - "Then you should worry"

Twist - Nuclear plant has two states

- heat and pressure mostly low
  - working condition
- Heat high and pressure varying wildly
  - "Then you should worry..."

When do you raise the alarm?

Motivation - "Then you should worry"

Israeltech challenge

Twist - Nuclear plant has two states

- heat and pressure mostly low
  - working condition
- Heat high and pressure varying wildly
  - "Then you should worry..."

When do you raise the alarm?

- You can use MLE
- You can derive a prior and use MAP

Motivation - When should you worry?

Original setting

- only one state is known - working condition
- can't decide which state (hypothesis) is more probable
- what can we do?

discuss

Motivation - When should you worry?

Original setting

- only one state is known - working condition
- can't decide which state (hypothesis) is more probable
- what can we do?
- we can raise the alarm when our (only) hypothesis becomes improbable

Null and Alternative Hypotheses

null hypothesis H_0

the hypothesis we test for improbability

alternative hypothesis H_1

- competing hypothesis
- frequently unknown

1 Introduction

2 Testing

- Common Errors
- Multiple Testing

3 Conclusion

The Night is Dark and Full of Error

Sitting at the control room, late at night, what are you afraid of?

discuss

The Night is Dark and Full of Error

Israeltech challenge

Sitting at the control room, late at night, what are you afraid of?

- Not raising an alarm on time - possible catastrophe
- Raising an alarm for nothing - the boy who cried wolf

The Night is Dark and Full of Error

Israeltech challenge

Sitting at the control room, late at night, what are you afraid of?

name	in our scenario	description
false negative	not raising an alarm on time	failure to detect an event/anomaly
false positive	raising an alarm for nothing	detecting an event/anomaly falsely

Mythbusters

A self proclaimed clairvoyant

Mythbusters

A self proclaimed clairvoyant

- Claim - able to predict outcome of coin tosses
- test with fair coin $(p = 0.5)$
- toss coin 6 times
- Null hypothesis - prediction is random

Mythbusters

A self proclaimed clairvoyant

- Claim - able to predict outcome of coin tosses
- test with fair coin $(p = 0.5)$
- toss coin 6 times
- Null hypothesis - prediction is random
- predicts 6/6
- probability of this event - $P(X \mid H_0) = 0.5^6$
- do you accept his claim?

probability, accept

Mythbusters

A self proclaimed clairvoyant

- Claim - able to predict outcome of coin tosses
- test with fair coin $(p = 0.5)$
- toss coin 6 times
- Null hypothesis - prediction is random
- predicts 5/6
- probability of event $P(X \mid H_0) = 0.5^6$

Mythbusters

A self proclaimed clairvoyant

- Claim - able to predict outcome of coin tosses
- test with fair coin $(p = 0.5)$
- toss coin 6 times
- Null hypothesis - prediction is random
- predicts 5/6
- probability of event $P(X \mid H_0) = \binom{6}{1} 0.5^6$
- do you accept his claim?

Mythbusters

Israeltech challenge

A self proclaimed clairvoyant

- predicts 5/6
- probability of event $P\left(X \mid H_0\right) = \binom{6}{1} 0.5^6$
- do you accept his claim?

What's you probability of wrongfully accepting?

- If you would accept for 5/6 you would also for 6/6

$$P_{FP} = P\left(6/6 \mid H_0\right) + P\left(5/6 \mid H_0\right)$$

P-Value

Definition

The probability of seeing a result at least as extreme as the one observed

---

- the probability of being wrong when **rejecting** the null hypothesis
- $Pval = P_{FP}$

The Correct Procedure

1. Determine a significance level α
   - maximal allowed False positive probability
2. Choose a Null hypothesis
3. Perform experiment
4. Calculate P-Value
5. Reject null hypothesis **if and only** if $Pval < \alpha$

Why Reject

1. Determine a significance level α
   - maximal allowed False positive probability
2. Choose a Null hypothesis
3. Perform experiment
4. Calculate P-Value
5. Reject null hypothesis **if and only** if $Pval < \alpha$

can't we also accept?

Why Reject

IsraelTech challenge

Coin inspector

- coin suspected of $p > 0.5$
- choose $\alpha = 0.1$
- H_0 -

Why Reject

Coin inspector

- coin suspected of $p > 0.5$
- choose $\alpha = 0.1$
- H_0 - coin is fair

Why Reject

Israeltech challenge

Coin inspector

- coin suspected of $p > 0.5$
- choose $\alpha = 0.1$
- H_0 - coin is fair
- Toss a coin 3 times - 2 heads
- $Pval = \binom{3}{1}0.5^3 + 0.5^3 = 0.5$
- Ignoring advice you accept the null hypothesis

Why Reject

Coin inspector

- coin suspected of $p > 0.5$
- choose $\alpha = 0.1$
- H_0 - coin is fair
- Toss a coin 3 times - 2 heads
- Your boss tosses coin 4 more times - 4 heads
- Overall $Pval = \binom{7}{1}0.5^7 + 0.5^7 = 0.0625$

Why Reject

You Don't Know What You Dont Know

- $Pval > \alpha$
  - result not unlikely given H_0
  - $P\left(X \mid H_0\right)$ not very small

You Don't Know What You Dont Know

- $Pval > \alpha$
  - result not unlikely given H_0
  - $P(X \mid H_0)$ not very small
- we know nothing about $P(H_0 \mid X)$
- we don't know the alternative hypothesis H_1

$$\text{possible} \begin{cases} P(X \mid H_1) > P(X \mid H_0) \\ P(H_1 \mid X) > P(H_0 \mid X) \end{cases}$$

You Don't Know What You Dont Know

IsraelTech challenge

In the previous example:

- H_1 may be $p = \dfrac{2}{3}$

$$P\left(X \geq 2 \mid H_1\right) = \binom{3}{1}\frac{2}{3}^2 \cdot \frac{1}{3} + \frac{2}{3}^3$$
$$\approxeq 0.741$$

$$P\left(X \geq 2 \mid H_0\right) = 0.5$$

Misstating The Null hypothesis

Disease - pineapple pen induced headaches

- Duration - almost always two weeks
- You're testing a new cure - plugear
- H_0 - plugear doesn't work
- α - 0.05
- A large test group is given plugear

Misstating The Null hypothesis

Disease - pineapple pen induced headaches

- 40% show major improvement in 1 day
- Only 0.0005 in the population show such an improvement
- Is plugear any good?

Misstating The Null hypothesis

Disease - pineapple pen induced headaches

- 40% show major improvement in 1 day
- Only 0.0005 in the population show such an improvement
- Is plugear any good?

Not necessarily

Misstating The Null hypothesis

Disease - pineapple pen induced headaches

- Why did our procedure fail?

Misstating The Null hypothesis

Disease - pineapple pen induced headaches

- Why did our procedure fail?
- Misstated null hypothesis
- No correction for a intrinsic bias

Misstating The Null hypothesis

The Placebo Effect

- Considering the effect
- $H_0 =$ plugear is no better than a placebo

experimental design

A coin inspector once more

- You are rehired
- task - find coins with $p > 0.5$
  - Choose $\alpha = 0.01$
  - Null hypothesis for a coin - coin is fair $p = 0.5$
  - You only reject
- You test 1000 coins

A coin inspector once more

- You find a coin with $Pval = 0.001 < \alpha = 0.01$
- Reject?

A coin inspector once more

- You find a coin with $Pval = 0.001 < \alpha = 0.01$
- Reject?

Not necessarily

A coin inspector once more

- event A - $Pval \geq 0.01$
- event $\overline{A}$ - $Pval < 0.01$

$$
\begin{aligned}
P\left(A \mid H_0\right) &= 1 - P\left(\overline{A} \mid H_0\right) \\
&= 0.99
\end{aligned}
$$

A coin inspector once more

you tested **1000** coins

- event B - $Pval < 0.01$ for $\boxed{\text{at least}}$ one coin
- event $\overline{B}$ - $Pval \geq 0.01$ for $\boxed{\text{all}}$ coins

$$\begin{aligned} P\left(B \mid H_0\right) &= 1 - P\left(\overline{B} \mid H_0\right) \\ &= 1 - 0.99^{1000} \\ &> 0.9999 \end{aligned}$$

Multiple Testing

Israeltech challenge

Setting

- m experiments
- $\forall i = 1, \ldots, m$ test H_0^i against unknown H_1^i

	H_0 not rejected	H_0 rejected	Total
H_0 is true	U	V	m_0
H_1 is true	T	S	$m - m_0$
Total	$m - R$	R	m

Bonferroni Correction

Procedure

- m experiments
- P_i P-Value for the i^{th} experiment
- reject H_0^i if $P_i < \dfrac{\alpha}{m}$

Bonferroni Correction - Properties

- Controls the $\boxed{\text{Family Wise Error Rate}}$
- event A - for $\boxed{\text{at least}}$ one i H_0^i is falsely rejected

Bonferroni Correction - Properties

- Controls the $\boxed{\text{Family Wise Error Rate}}$
- event A - for $\boxed{\text{at least}}$ one i H_0^i is falsely rejected

$$
\begin{aligned}
P(A) &= P\left(\bigcup_{i=1}^{m_0}\left\{P_i \leq \frac{\alpha}{m}\right\}\right) \\
\text{[union bound]} &\leq \sum_{i=1}^{m_0} P\left(P_i \leq \frac{\alpha}{m}\right) \\
&\leq m_0 \frac{\alpha}{m} \\
&\leq \alpha
\end{aligned}
$$

Bonferroni Correction - Properties

- no assumptions about dependence
- extremely conservative
  - coin example
  - per coin $Pval < 0.00001$
  - increased rate of false negatives

False Discovery Rate

Control Procedure

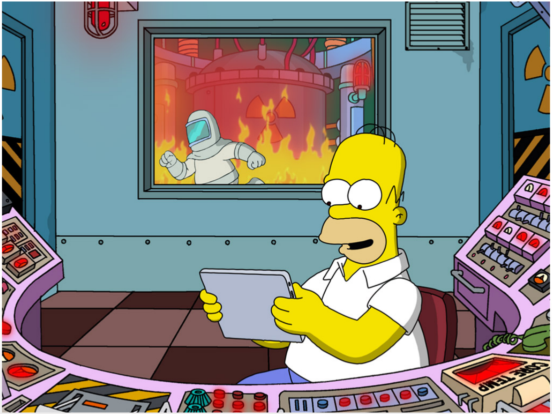
- a mission to mars
  - Family Wise Error Rate (FWER)
- manufacturing, surveillance
  - FWER is too strict
  - you must allow some false positives
  - why?

False Discovery Rate

In some cases it's about proportion

False Discovery Rate

In some cases it's about proportion



False Discovery Rate

you work 1000 shifts (experiments)

- raise 10 false alarms out of 11 alarms ...

False Discovery Rate

you work 1000 shifts (experiments)

- raise 10 false alarms out of 11 alarms ...
- raise 10 false alarms out of 200 alarms

Multiple Testing - FDR

Israeltech challenge

	H_0 not rejected	H_0 rejected	Total
H_0 is true	U	V	m_0
H_1 is true	T	S	$m - m_0$
Total	$m - R$	R	m

- $\boxed{Q = \frac{V}{R}}$ - proportion of false discoveries $R = 0$
- $\boxed{FDR = E\left[Q\right]}$ - false discovery rate

Benjamini-Hochberg Method

IsraelTech challenge

- order (rank) P-Values for all tests

$$P_{(1)} \leq \ldots \leq P_{(m)}$$

- define threshold for each (i)

$$l_{(i)} = \frac{k}{m \cdot c\left(m\right)}\alpha$$

  - independent tests $c\left(m\right) = 1$
  - dependence $c\left(m\right) = \sum_{i=1}^{m} \frac{1}{i}$

Yekutieli

Benjamini-Hochberg Method

Decision

- find largest k such that $P_{(k)} \leq l_{(k)}$
- reject null hypothesis for $P_{(1)}, \ldots, P_{(k)}$

Under independence and some forms of dependence

$$\boxed{FDR_{BH} \leq \frac{m_0}{m}\alpha \leq \alpha}$$

Discover, detect <u>but</u> don't over do it

- a new kind of statistical reasoning
- how do you say surprise in statistics
- how not to be a bad (data) scientist
- multiple testing

exercise - batch

Credits

figures

- Homer at work - feelgood.network/author/cattownsend/
- Pressure dial - commons.wikimedia.org/wiki/File:Psidial.jpg
- Clairvoyant - www.quickmeme.com/baby-psychic