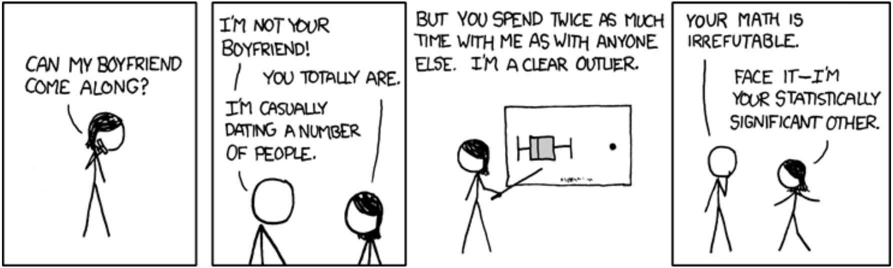
Frequentist Statistics and Hypothesis Testing

18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom



http://xkcd.com/539/

Agenda

- Introduction to the frequentist way of life.
- What is a statistic?

- NHST ingredients; rejection regions
- Simple and composite hypotheses
- z-tests, p-values

Frequentist school of statistics

- Dominant school of statistics in the 20^{th} century.
- *p*-values, *t*-tests, χ^2-tests, confidence intervals.
- Defines probability as long-term frequency in a repeatable random experiment.
  - Yes: probability a coin lands heads.
  - Yes: probability a given treatment cures a certain disease.
  - Yes: probability distribution for the error of a measurement.
- Rejects the use of probability to quantify incomplete knowledge, measure degree of belief in hypotheses.
  - No: prior probability for the probability an unknown coin lands heads.
  - No: prior probability on the efficacy of a treatment for a disease.
  - No: prior probability distribution for the unknown mean of a normal distribution.

The fork in the road

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Everyone uses Bayes' formula when the prior is known.

Bayesian path

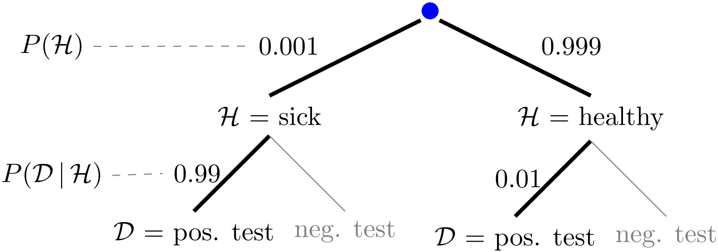
Frequentist path

Bayesians require a prior, so they develop one from the best information they have.

Frequentists reject the prior and draw inferences as best they can from just the likelihood function $P(D|H)$.

Disease screening redux: probability

The test is positive. Are you sick?

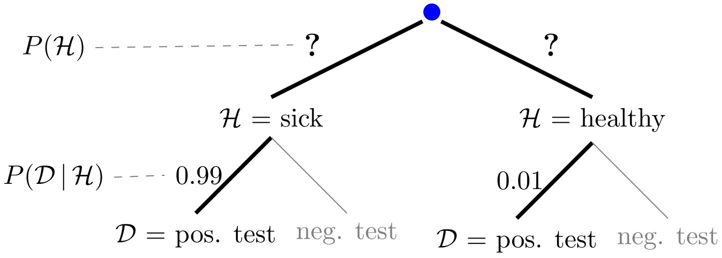


The prior is known so we can use Bayes Theorem.

$$P(\mathcal{H} = \text{ sick} \mid \mathcal{D}) = \frac{0.001 \cdot 0.99}{0.001 \cdot 0.99 + 0.999 \cdot 0.01} \approx 0.1$$

Disease screening redux: statistics

The test is positive. Are you sick?



The prior is not known.
Bayesian: use a subjective prior $P(\mathcal{H})$ and Bayes Theorem.

Frequentist: the likelihood is all we can use: $P(\mathcal{D} \mid \mathcal{H})$

Concept question

Each day Jane arrives X hours late to class, with $X \sim \text{uniform}(0, \theta)$. Jon models his initial belief about θ by a prior pdf $f(\theta)$. After Jane arrives x hours late to the next class, Jon computes the likelihood function $f(x|\theta)$ and the posterior pdf $f(\theta|x)$.

Which of these probability computations would the frequentist consider valid?

1. none
2. prior
3. likelihood
4. posterior
5. prior and posterior
6. prior and likelihood
7. likelihood and posterior
8. prior, likelihood and posterior.

Statistics are computed from data

Working definition. A *statistic* is anything that can be computed from random data.

A statistic *cannot* depend on the value of an unknown parameter.

Examples of point statistics

- Data mean
- Data maximum (or minimum)
- Maximum likelihood estimate (MLE)

A statistic is random since it is computed from random data.

We can also get more complicated statistics like **interval statistics.**

Concept questions

Suppose $x_1, \ldots, x_n$ is a sample from $N(\mu, \sigma^2)$, where μ and σ are unknown.

Is each of the following a statistic?

1. Yes 2. No

1. The median of $x_1, \ldots, x_n$.

Concept questions

Suppose $x_1, \ldots, x_n$ is a sample from $\mathrm{N}(\mu, \sigma^2)$, where μ and σ are unknown.

Is each of the following a statistic?

1. Yes 2. No

1. The median of $x_1, \ldots, x_n$.
2. The interval from the .25 quantile to the .75 quantile of $\mathrm{N}(\mu, \sigma^2)$.

Concept questions

Suppose $x_1, \ldots, x_n$ is a sample from $\mathrm{N}(\mu, \sigma^2)$, where μ and σ are unknown.

Is each of the following a statistic?

1. Yes 2. No

1. The median of $x_1, \ldots, x_n$.
2. The interval from the .25 quantile to the .75 quantile of $\mathrm{N}(\mu, \sigma^2)$.
3. The standardized mean $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

Concept questions

Suppose $x_1, \ldots, x_n$ is a sample from $N(\mu, \sigma^2)$, where μ and σ are unknown.

Is each of the following a statistic?

1. Yes 2. No

1. The median of $x_1, \ldots, x_n$.
2. The interval from the .25 quantile to the .75 quantile of $N(\mu, \sigma^2)$.
3. The standardized mean $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.
4. The set of sample values less than 1 unit from $\bar{x}$.

Cards and NHST

Illustration of a royal flush removed due to copyright restrictions.

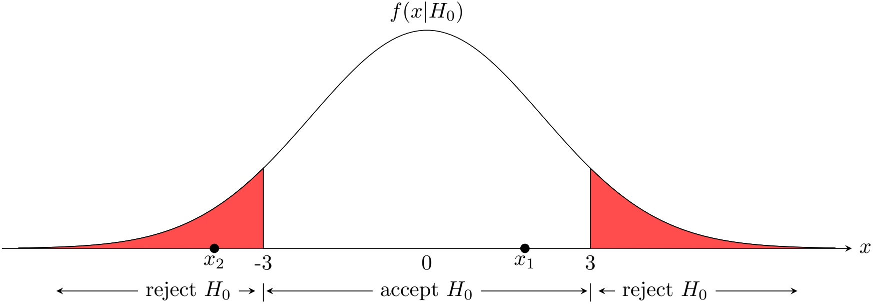
NHST ingredients

Null hypothesis: H_0

Alternative hypothesis: H_A

Test statistic: x

Rejection region: reject H_0 in favor of H_A if x is in this region



$p(x|H_0)$ or $f(x|H_0)$: null distribution

Choosing rejection regions

Coin with probability of heads θ.

Test statistic x = the number of heads in 10 tosses.

H_0: 'the coin is fair', i.e. $\theta = .5$

H_A: 'the coin is biased, i.e. $\theta \neq .5$

Two strategies:

1. Choose rejection region then compute significance level.

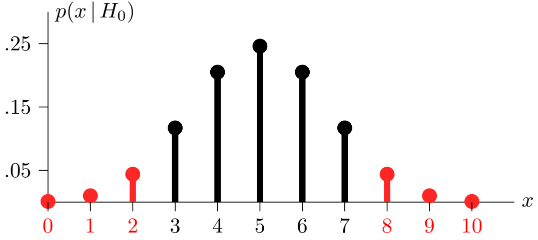
2. Choose significance level then determine rejection region.

***** Everything is computed assuming H_0 *****

Board question

Suppose we have the coin from the previous slide.

1. The rejection region is bordered in red, what's the significance level?



x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

2. Given significance level $\alpha = .05$ find a two-sided rejection region.

Solution

1. $\alpha = .11$

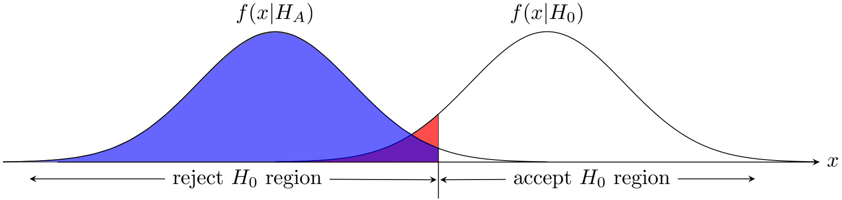
x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

2. $\alpha = .05$

x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

Concept question

The null and alternate pdfs are shown on the following plot



The significance level of the test is given by the area of which region?

1. red   2. purple   3. blue   4. white
5. blue + purple   6. red + purple   7. white + red + purple.

z-tests, p-values

Normal Data: $x_1, \ldots, x_n$; unknown mean μ, known σ

Hypotheses: H_0: $x_i \sim N(\mu_0, \sigma^2)$

H_A: Two-sided: $\mu \neq \mu_0$, or one-sided: $\mu > \mu_0$

Test statistic: $\bar{x}$

Null distribution: $\bar{x} \sim \mathrm{N}(\mu_0, \sigma^2/n)$.

z-**value:** standardized $\bar{x}$: $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$

p-**values:** Two-sided p-**value:** $p = P(|Z| > z \mid H_0)$

Right-sided p-**value:** $p = P(Z > z \mid H_0)$

Significance level: For $p \leq \alpha$ we reject H_0 in favor of H_A.

Visualization

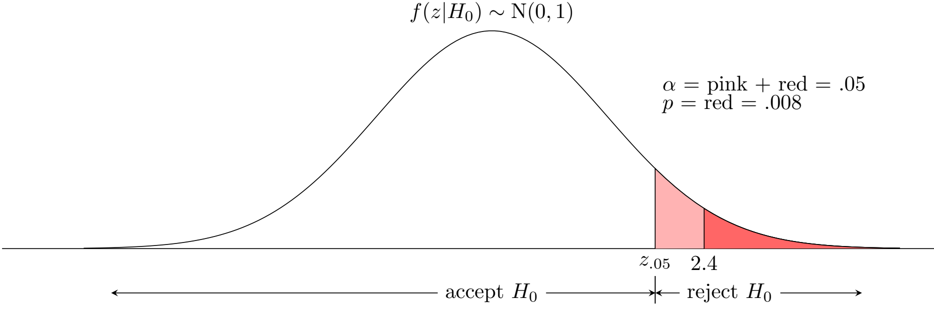
Data follows a normal distribution $\mathrm{N}(\mu, 15^2)$ where μ is unkown.

H_0: $\mu = 100$

H_A: $\mu > 100$ (one-sided)

Collect 9 data points: $\bar{x} = 112$

Can we reject H_0 at significance level .05?



Board question

- H_0: data follows a $N(5, 10^2)$
- H_A: data follows a $N(\mu, 10^2)$ where $\mu \neq 5$.
- Test statistic: $\overline{x}$ the average of the data.
- Data: 64 data points with $\overline{x} = 6.25$.
- Significance level set to $\alpha = .05$.

(i) Find the rejection region; draw a picture.

(ii) Find the z-value.

(iii) Decide whether or not to reject H_0 in favor of H_A.

(iv) Find the p-value for this data; add to your picture.

Board question

Two coins: probability of heads is .5 for C_1; and .6 for C_2.

We pick one at random, flip it 8 times and get 6 heads.

1. $H_0 =$ 'The coin is C_1' $H_A =$ 'The coin is C_2'

Do you reject H_0 at the significance level $\alpha = .05$?

2. $H_0 =$ 'The coin is C_2' $H_A =$ 'The coin is C_1'

Do you reject H_0 at the significance level $\alpha = .05$?

3. Do your answers to (1) and (2) seem paradoxical?

Here are binomial(8,θ) tables for $\theta = .5$ and .6.

k	0	1	2	3	4	5	6	7	8
$p(k\|\theta = .5)$	.004	.031	.109	.219	.273	.219	.109	.031	.004
$p(k\|\theta = .6)$	.001	.008	.041	.124	.232	.279	.209	.090	.017