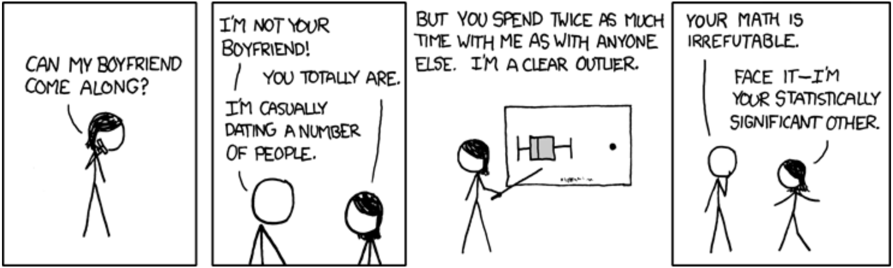
# Frequentist Statistics and Hypothesis Testing

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, $\chi^2$-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

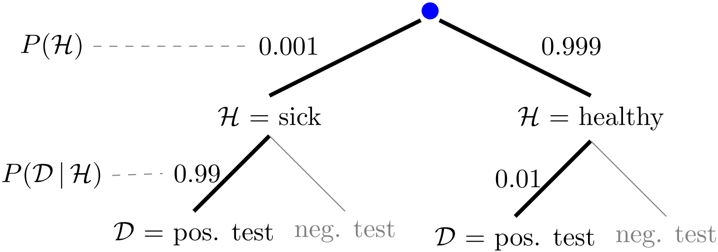
Bayesians require a prior, so they develop one from the best information they have.

Frequentist path

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?

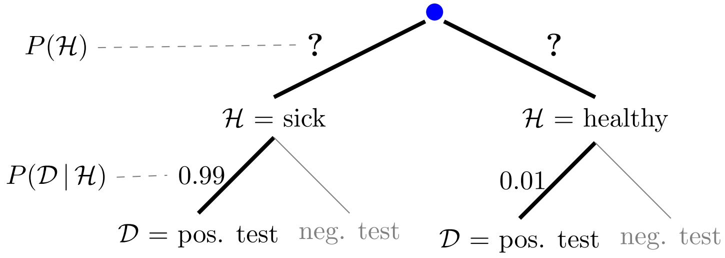



The prior is known so we can use Bayes Theorem.

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

# 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 $\theta$ 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.

# Concept answer

<u>**answer:**</u> 3. likelihood

Both the prior and posterior measure a belief in the distribution of hypotheses about the value of $\theta$. The frequentist does not consider them valid.

The likelihood $f(x|theta)$ is perfectly acceptable to the frequentist. It represents the probability of data from a repeatable experiment, i.e. measuring how late Jane is each day. Conditioning on $\theta$ is fine. This just fixes a model parameter $\theta$. It doesn't require computing probabilities for values of $\theta$.

# 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 $\mathrm{N}(\mu, \sigma^2)$, where $\mu$ and $\sigma$ 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 $\mu$ and $\sigma$ 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 $N(\mu, \sigma^2)$, where $\mu$ and $\sigma$ 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 $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

# Concept questions

Suppose $x_1, \ldots, x_n$ is a sample from $N(\mu, \sigma^2)$, where $\mu$ and $\sigma$ 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 $\bar{x}$.

# Concept answers

1. Yes. The median only depends on the data $x_1, \ldots, x_n$.

2. No. This interval depends only on the distribution parameters $\mu$ and $\sigma$. It does not consider the data at all.

3. No. this depends on the values of the unknown parameters $\mu$ and $\sigma$.

4. Yes. $\bar{x}$ depends only on the data, so the set of values within 1 of $\bar{x}$ can all be found by working with the data.

# Cards and NHST

Illustration of a royal flush removed due to copyright restrictions.

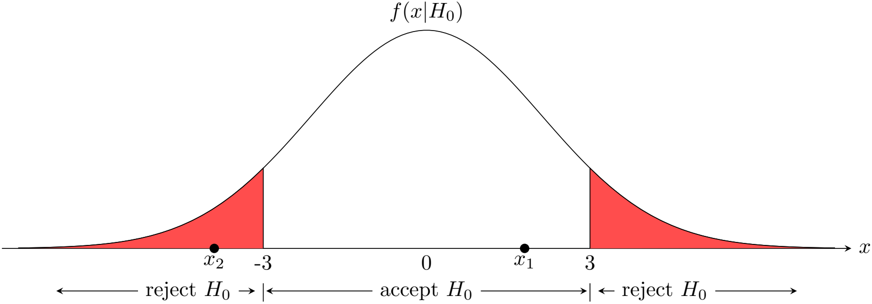
# NHST ingredients

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 $\theta$.

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.

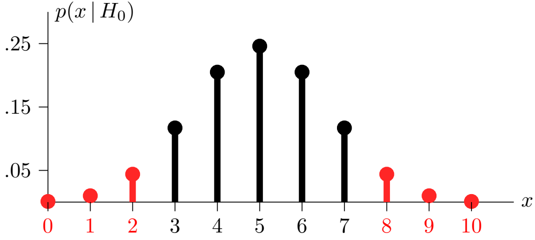
2. Choose significance level then determine rejection region.

***** 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.** $\alpha = .11$

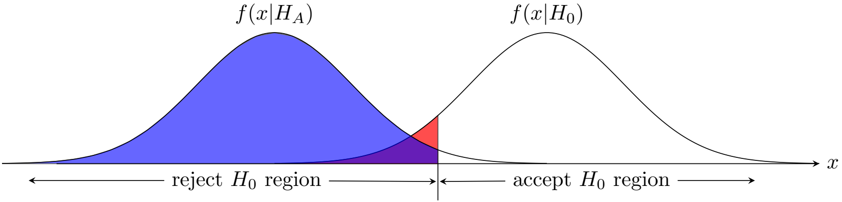
| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p(x\|H_0)$ | .001 | .010 | .044 | .117 | .205 | .246 | .205 | .117 | .044 | .010 | .001 |

**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.

<u>**answer:**</u> 6. Red + purple. This is the area under the pdf for $H_0$ above the rejection region.

# z-tests, p-values

**Normal Data:** $x_1, \ldots, x_n$; unknown mean $\mu$, known $\sigma$

**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

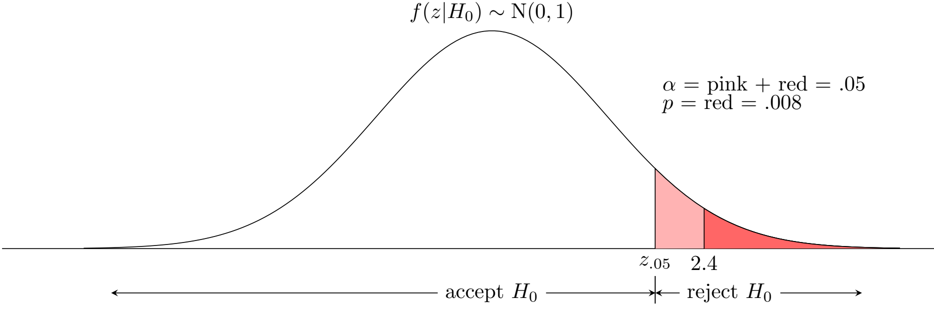
Data follows a normal distribution $N(\mu, 15^2)$ where $\mu$ is unkown.

$H_0$: $\mu = 100$

$H_A$: $\mu > 100$ (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.

**answer:** *See next slide*

# Solution

The null distribution $f(\overline{x} \mid H_0) \sim N(5, (10/8)^2)$

(i) The rejection region is $|\overline{x} - 5| > 5/2$, i.e. $(-\infty, 2.5) \cup (7.5, \infty)$, i.e more than two standard deviations from the mean.

(ii) Standardizing $z = \dfrac{\overline{x} - 5}{5/4} = \dfrac{1.25}{1.25} = 1.$

(iii) Do not reject since $z$ is not in the rejection region

(iv) Use a two-sided $p$-value $p = P(|Z| > 1) = .32$

Must add figure

# 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,$\theta$) 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 |

# Solution

**1.** Since $.6 > .5$ we use a right-sided rejection region.
Under $H_0$ the probability of heads is .5. Using the table we find a one sided rejection region $\{7, 8\}$. That is we will reject $H_0$ in favor of $H_A$ only if we get 7 or 8 heads in 8 tosses.
Since the value of our data $x = 6$ is not in our rejection region we do not reject $H_0$.

**2.** Since $.6 > .5$ we use a left-sided rejection region.
Now under $H_0$ the probability of heads is .6. Using the table we find a one sided rejection region $\{0, 1, 2\}$. That is we will reject $H_0$ in favor of $H_A$ only if we get 0, 1 or 2 heads in 8 tosses.
Since the value of our data $x = 6$ is not in our rejection region we do not reject $H_0$.