Frequentist Statistics and Hypothesis Testing
18.05 Spring 2014

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 20th 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

Bayesian path

\[
P(H|D) = \frac{P(D|H)P(H)}{P(D)}
\]

Everyone uses Bayes’ formula when the prior \( P(H) \) is known.

Frequentist path

\[
P_{\text{Posterior}}(H|D) = \frac{P(D|H)P_{\text{prior}}(H)}{P(D)}
\]

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

\[
L(H; D) = P(D|H)
\]

Without a known prior frequentists draw inferences from just the likelihood function.
Disease screening redux: probability

The test is positive. Are you sick?

\[
P(\text{sick} \mid \text{pos. test}) = \frac{0.001 \cdot 0.99}{0.001 \cdot 0.99 + 0.999 \cdot 0.01} \approx 0.1
\]

The prior is known so we can use Bayes’ Theorem.
The test is positive. Are you sick?

\[
P(\mathcal{H}) \quad \quad \quad \quad ?
\]

\[\mathcal{H} = \text{sick} \quad \quad \quad \quad \quad \quad \quad \quad ?\]

\[
P(D | \mathcal{H}) \quad \quad \quad \quad 0.99
\]

\[D = \text{pos. test} \quad \quad \text{neg. test} \]

\[
P(D | \mathcal{H}) \quad \quad \quad \quad 0.01
\]

\[D = \text{pos. test} \quad \text{neg. test} \]

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

Frequentist: the likelihood is all we can use: \( P(D | \mathcal{H}) \)
Each day Jane arrives $X$ hours late to class, with $X \sim \text{uniform}(0, \theta)$, where $\theta$ is unknown. 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.
Statistics are computed from data

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

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

A statistic **can** depend on a hypothesized value of a 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 $\mu$ 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 0.25 quantile to the 0.75 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
**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. $\theta = 0.5$

$H_A$: ‘the coin is biased, i.e. $\theta \neq 0.5$

**Two strategies:**

1. Choose rejection region then compute significance level.

2. Choose significance level then determine rejection region.

***** Everything is computed assuming $H_0$ *****
Table question

Suppose we have the coin from the previous slide.

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

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

1. $\alpha = 0.11$

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x</td>
<td>H_0)$</td>
<td>.001</td>
<td>.010</td>
<td>.044</td>
<td>.117</td>
<td>.205</td>
<td>.246</td>
<td>.205</td>
<td>.117</td>
<td>.044</td>
<td>.010</td>
</tr>
</tbody>
</table>

2. $\alpha = 0.05$

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x</td>
<td>H_0)$</td>
<td>.001</td>
<td>.010</td>
<td>.044</td>
<td>.117</td>
<td>.205</td>
<td>.246</td>
<td>.205</td>
<td>.117</td>
<td>.044</td>
<td>.010</td>
</tr>
</tbody>
</table>
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. \( R_1 \)
2. \( R_2 \)
3. \( R_3 \)
4. \( R_4 \)
5. \( R_1 + R_2 \)
6. \( R_2 + R_3 \)
7. \( R_2 + R_3 + R_4 \).
z-tests, p-values

Suppose we have independent normal Data: \( x_1, \ldots, x_n \); with 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: \( \mu > \mu_0 \)

**z-value:**

standardized \( \bar{x} \): \( z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \)

**Test statistic:**

\( z \)

**Null distribution:**

Assuming \( H_0: z \sim N(0, 1) \).

**p-values:**

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

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

**Significance level:**

For \( p \leq \alpha \) we reject \( H_0 \) in favor of \( H_A \).

Note: Could have used \( \bar{x} \) as test statistic and \( N(\mu_0, \sigma^2) \) as the null distribution.
Visualization

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

$H_0$: $\mu = 100$

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

Collect 9 data points: $\bar{x} = 112$. So, $z = \frac{112 - 100}{15/3} = 2.4$.

Can we reject $H_0$ at significance level 0.05?

Can we reject $H_0$ at significance level 0.05?

$z_{0.05} = 1.64$

$\alpha = \text{pink} + \text{red} = 0.05$

$p = \text{red} = 0.008$
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: $z = \text{standardized } \bar{x}$.
- Data: 64 data points with $\bar{x} = 6.25$.
- Significance level set to $\alpha = 0.05$.

(i) Find the rejection region; draw a picture.
(ii) Find the z-value; add it to your picture.
(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.
(v) What’s the connection between the answers to (ii), (iii) and (iv).
Solution

The null distribution \( f(z \mid H_0) \sim N(0, 1) \)

(i) The rejection region is \(|z| > 1.96\), i.e. 1.96 or more standard deviations from the mean.

(ii) Standardizing \( z = \frac{\bar{x} - 5}{5/4} = \frac{1.25}{1.25} = 1 \).

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

(iv) Use a two-sided \( p \)-value \( p = P(|Z| > 1) = .32 \)
(v) The $z$-value is not in the rejection region tells us exactly the same thing as the $p$-value being greater than the significance, i.e. don’t reject the null hypothesis $H_0$. 
Board question

Two coins: probability of heads is 0.5 for $C_1$; and 0.6 for $C_2$.

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

1. $H_0 = \text{'The coin is } C_1\text{' }$ $H_A = \text{'The coin is } C_2\text{' }$
   
   Do you reject $H_0$ at the significance level $\alpha = 0.05$?

2. $H_0 = \text{'The coin is } C_2\text{' }$ $H_A = \text{'The coin is } C_1\text{' }$
   
   Do you reject $H_0$ at the significance level $\alpha = 0.05$?

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

Here are binomial$(8, \theta)$ tables for $\theta = 0.5$ and 0.6.

| k | $p(k|\theta = 0.5)$ | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  |
|---|---------------------|----|----|----|----|----|----|----|----|----|
|   |                     | 004| 031| 109| 219| 273| 219| 109| 031| 004|
|   | $p(k|\theta = 0.6)$ | .001| .008| .041| .124| .232| .279| .209| .090| .017|
18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.