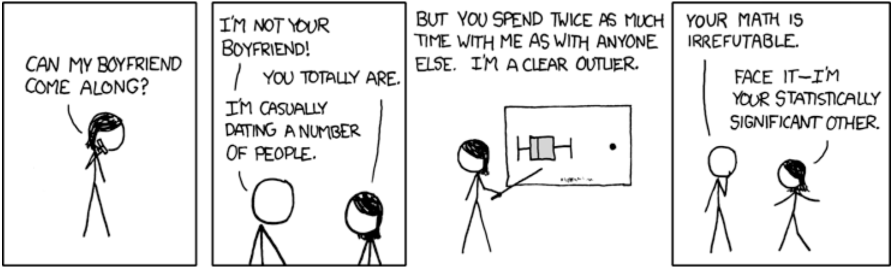
# Frequentist Statistics and Hypothesis Testing
# 18.05 Spring 2022






https://xkcd.com/539/

# Announcements/Agenda

**Announcements**

- None

**Agenda**

- New topic: many terms, many concept questions.
- Pay attention to the pictures, both shading and labels.

- 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 $20^{\text{th}}$ 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 efficac of a treatment for a disease.
  - No: prior probability distribution for the unknown mean of a normal distribution.

# The fork in the road

**Probability (mathematics)**

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

Everyone uses Bayes' formula when the prior $P(H)$ is known.

Bayesian path

Frequentist path

**Statistics (art)**

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

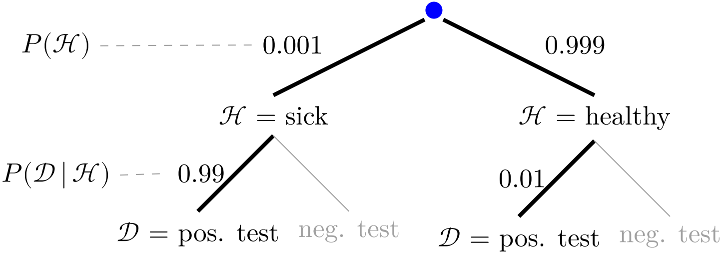
Bayesians require a prior, so they develop one from the best information they have.

Likelihood $L(H;D) = P(D|H)$

Without a known prior, frequentists draw inferences from just the likelihood function.

# Disease screening redux: probability

The test is positive. Are you sick?

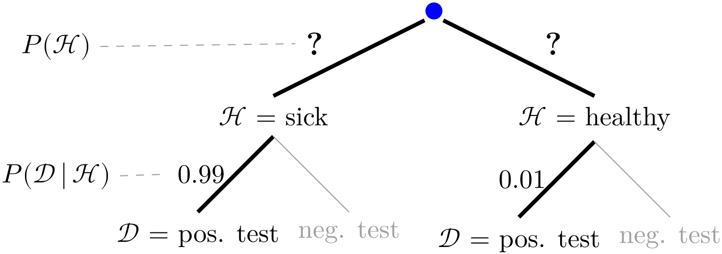



The prior is known so we can use Bayes' Theorem.

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

# 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: What would a frequentist say?

Each day Jaam 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 Jaam arrives $x$ hours late to the next class, Jon computes the likelihood function $\phi(x|\theta)$ 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 and known values.

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.**

# Discussion questions: Is it a statistic?

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

Is each of the following a statistic?

**(a)** The median of $x_1, \ldots, x_n$.

# Discussion questions: Is it a statistic?

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

Is each of the following a statistic?

**(a)** The median of $x_1, \ldots, x_n$.

**(b)** The interval from the 0.25 quantile to the 0.75 quantile of $\mathsf{N}(\mu, \sigma^2)$.

# Discussion questions: Is it a statistic?

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

Is each of the following a statistic?

**(a)** The median of $x_1, \dots, x_n$.

**(b)** The interval from the 0.25 quantile to the 0.75 quantile of $\mathsf{N}(\mu, \sigma^2)$.

**(c)** The standardized mean $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.

# Discussion questions: Is it a statistic?

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

Is each of the following a statistic?

**(a)** The median of $x_1, \dots, x_n$.

**(b)** The interval from the 0.25 quantile to the 0.75 quantile of $\mathsf{N}(\mu, \sigma^2)$.

**(c)** The standardized mean $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.

**(d)** The set of sample values less than 1 unit from $\bar{x}$.

# Discussion questions: Is it a statistic?

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

Is each of the following a statistic?

**(a)** The median of $x_1, \dots, x_n$.

**(b)** The interval from the 0.25 quantile to the 0.75 quantile of $\mathrm{N}(\mu, \sigma^2)$.

**(c)** The standardized mean $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.

**(d)** The set of sample values less than 1 unit from $\bar{x}$.

**(e)** The $z = \dfrac{\overline{x} - 5}{3/\sqrt{n}}$.

# Discussion questions: Is it a statistic?

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

Is each of the following a statistic?

**(a)** The median of $x_1, \dots, x_n$.

**(b)** The interval from the 0.25 quantile to the 0.75 quantile of $\mathsf{N}(\mu, \sigma^2)$.

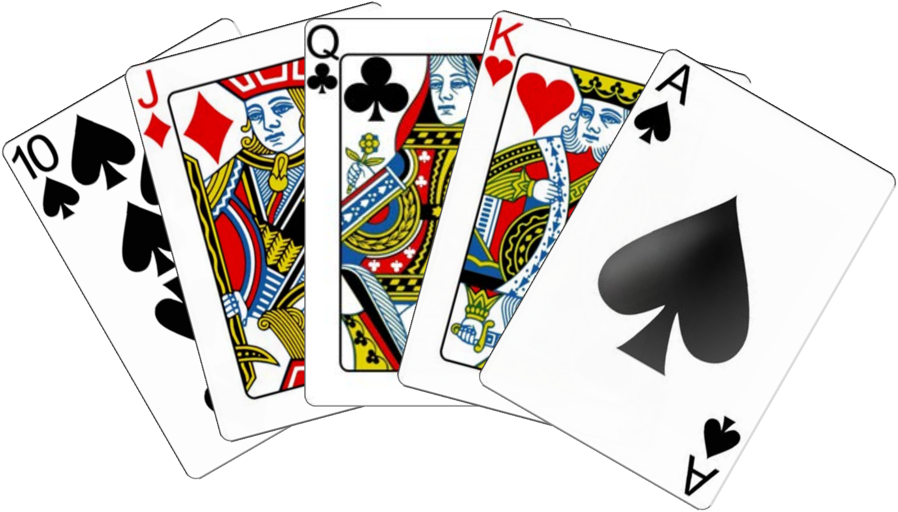
**(c)** The standardized mean $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.

**(d)** The set of sample values less than 1 unit from $\bar{x}$.

**(e)** The $z = \dfrac{\overline{x} - 5}{3/\sqrt{n}}$.

**(f)** $z = \dfrac{\overline{x} - \mu_0}{\sigma_0/\sqrt{n}}$, where $\mu_0$ and $\sigma_0$ are given values,

# Cards and NHST

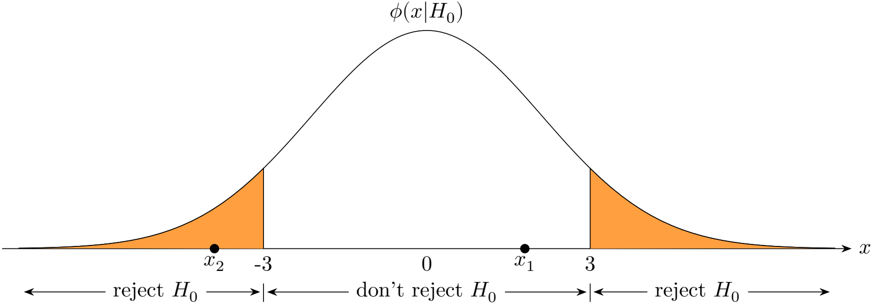


# NHST ingredients

Null hypothesis: $H_0$, Alternative hypothesis: $H_A$

Test statistic: $x$ (computed from the data)

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




$p(x|H_0)$ or $\phi(x|H_0)$: null distribution

We define **significance** $\alpha = P(x \text{ is in the rejection region} \mid H_0)$.

# 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: Testing coins

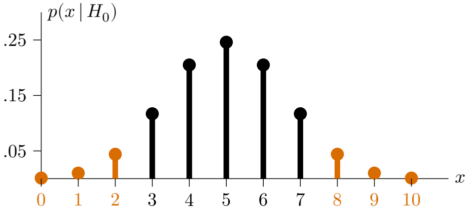
Suppose we have a coin with an unknown probability of heads $\theta$.

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

Null hypothesis $H_0$: $\theta = 0.5$ (fair coin).

Alternative hypothesis $H_A$: $\theta \neq 0.5$ (unfair coin, two-sided).

**(a)** The rejection region is are the values of $x$ shown in orange. What's the significanc level?




| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p(x\|H_0)$ | 0.001 | 0.010 | 0.044 | 0.117 | 0.205 | 0.246 | 0.205 | 0.117 | 0.044 | 0.010 | 0.001 |

**(b)** For significanc level $\alpha = 0.05$, fin a two-sided rejection region.

# Solution

**(a)** $\alpha = 0.11$

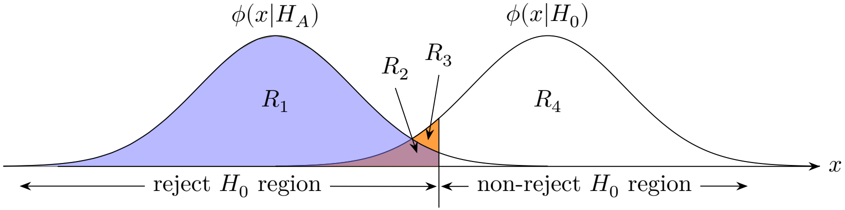
| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p(x\|H_0)$ | 0.001 | 0.010 | 0.044 | 0.117 | 0.205 | 0.246 | 0.205 | 0.117 | 0.044 | 0.010 | 0.001 |

**(b)** $\alpha = 0.05$

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p(x\|H_0)$ | 0.001 | 0.010 | 0.044 | 0.117 | 0.205 | 0.246 | 0.205 | 0.117 | 0.044 | 0.010 | 0.001 |

# Concept question: Picture the significance.

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. $R_3$
3. $R_1 + R_2$
4. $R_2 + R_3 + R_4$
5. $R_2$
6. $R_4$
7. $R_2 + R_3$
8. None of these

# z-tests, p-values (two-sided)

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: $\mu \neq \mu_0$. |
| $z$**-value:** | standardized $\overline{x}$: $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$ |
| **Test statistic:** | $z$ |
| **Null distribution:** | Assuming $H_0$: $z \sim \mathrm{N}(0, 1)$. |
| **Significance:** | Choose significanc level $\alpha$ (often $\alpha = 0.05$). |
| **Two-sided rejection region:** | Left and right tails, with probability $\alpha/2$ each. |
| **Two sided $p$ value:** | $p = P(\lvert Z \rvert > \lvert z \rvert \mid H_0)$ |
| **Test:** | For $p \leq \alpha$ we reject $H_0$ in favor of $H_A$, i.e. if $p \leq \alpha$, then $z$ is in the rejection region. |

# z-tests, p-values (right-sided)

Independent **normal Data:** $x_1, \ldots, x_n$; with unknown mean $\mu$, known $\sigma$

| | |
|---|---|
| **Hypotheses:** | $H_0$: $x_i \sim N(\mu_0, \sigma^2)$<br>$H_A$: Right-sided: $\mu > \mu_0$. |
| $z$**-value:** | standardized $\overline{x}$: $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$ |
| **Test statistic:** | $z$ |
| **Null distribution:** | Assuming $H_0$: $z \sim \mathrm{N}(0, 1)$. |
| **Significance:** | We set a significanc level $\alpha$ (often $\alpha = 0.05$). |
| **Right-sided rejection region:** | Right tail, with probability $\alpha$. |
| **Right-sided $p$ value:** | $p = P(Z > z \,\vert\, H_0)$ |
| **Test:** | For $p \le \alpha$ we reject $H_0$ in favor of $H_A$, i.e. if $p \le \alpha$, then $z$ is in the rejection region. |

# Visualization example

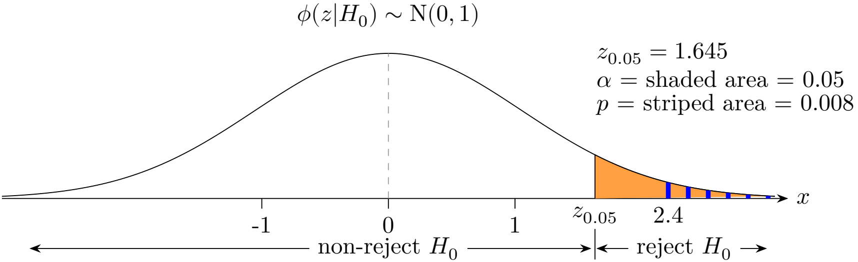
Suppose our data follows a normal distribution $\text{N}(\mu, 15^2)$, where $\mu$ is unknown.

Suppose our null and alternative hypotheses are

$$H_0\colon\ \mu = 100 \qquad H_A\colon\ \mu > 100 \text{ (one-sided)}$$

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

Can we reject $H_0$ at significance level 0.05?



# Board question: z statistic

Suppose we know the following about our null hypothesis significance test.

- $H_0$: data follows a $N(5, 10^2)$
- $H_A$: data follows a $N(\mu, 10^2)$ where $\mu \neq 5$.
- Test statistic: $z =$ standardized $\overline{x}$.
- Data: 64 data points with $\overline{x} = 6.25$.
- Significance level set to $\alpha = 0.05$.

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

**(b)** Find the $z$-value; add it to your picture.

**(c)** Decide whether or not to reject $H_0$ in favor of $H_A$.

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

**(e)** What's the connection between the answers to (b), (c) and (d)?

# Solution parts a-d

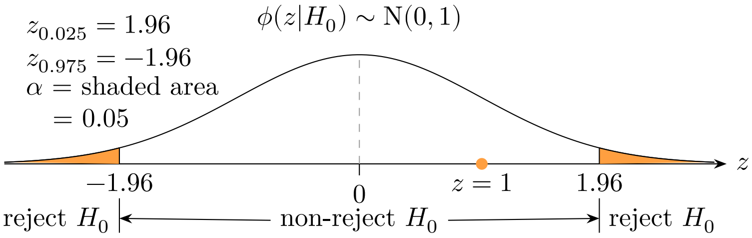
The null distribution $\phi(z \,|\, H_0) \sim N(0,1)$

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

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

**(c)** Do not reject since $z$ is not in the rejection region.

**(d)** Use a two-sided $p$-value $p = P(|Z| > 1) = 0.32.$



# Solution part e

**(e)** The $z$-value not being 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: More coins

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

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

Here are the probability tables for the two coins

| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| $p(k\|\theta = 0.5)$ | 0.004 | 0.031 | 0.109 | 0.219 | 0.273 | 0.219 | 0.109 | 0.031 | 0.004 |
| $p(k\|\theta = 0.6)$ | 0.001 | 0.008 | 0.041 | 0.124 | 0.232 | 0.279 | 0.209 | 0.090 | 0.017 |

**(a)** $H_0$ = 'The coin is $C_1$'  $H_A$ = 'The coin is $C_2$'

Do you reject $H_0$ at the significanc  level $\alpha = 0.05$?

(Hint: First decide if this test is two-sided, left-sided or right-sided. Then determine the rejection region.)

**(b)** $H_0$ = 'The coin is $C_2$'  $H_A$ = 'The coin is $C_1$'

Do you reject $H_0$ at the significanc  level $\alpha = 0.05$?

**(c)** Do your answers to (a) and (b) seem paradoxical