Review session

Intro Statistics: Life Sciences

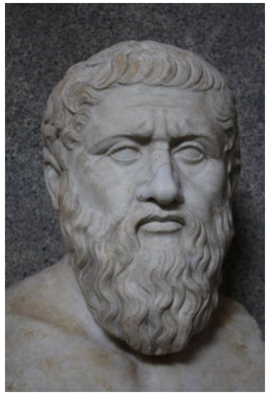
Descriptive and inferential statistics

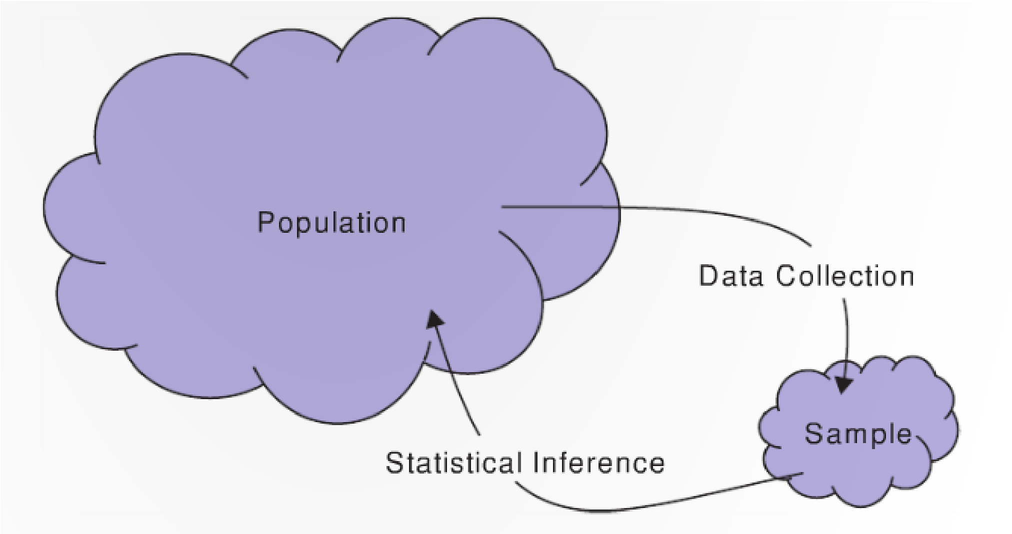
Descriptive Statistics: describe the sample of data we have

- i.e., describe the shadows

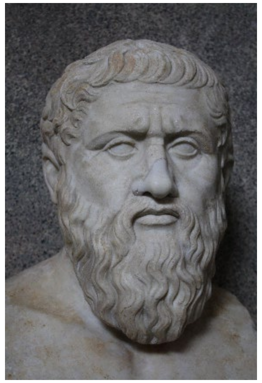
Inferential Statistics: use the sample to make claims about properties of the population/process

- i.e., try to use the data to get at the Truth

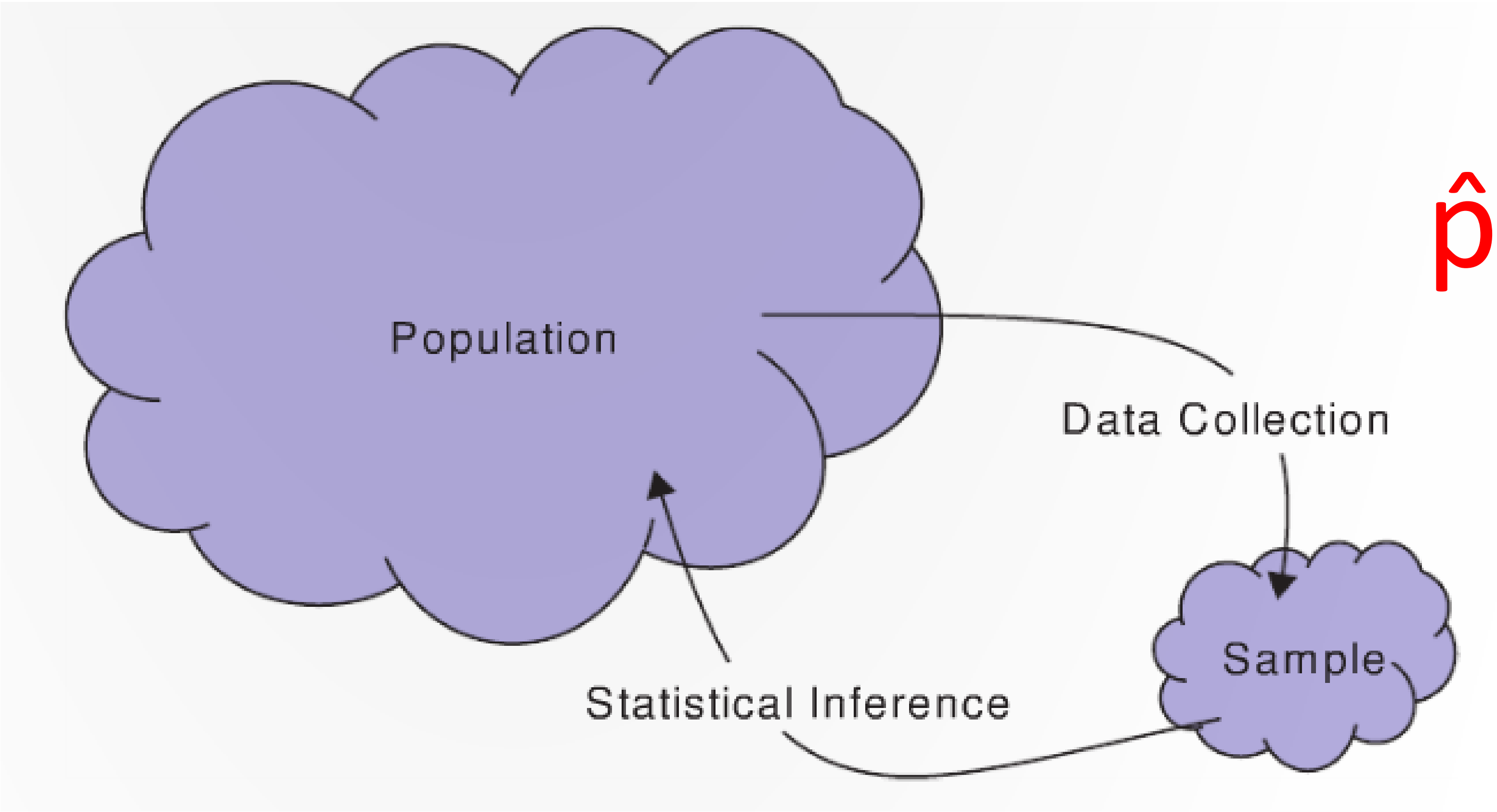




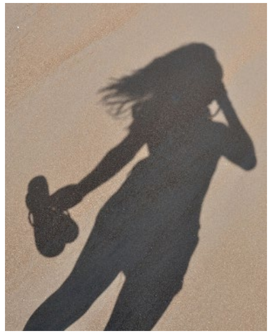
$\pi, \mu, \sigma, \rho, \beta$



Population: all individuals/objects of interest



$\hat{p}, \bar{x}, s, r, b$

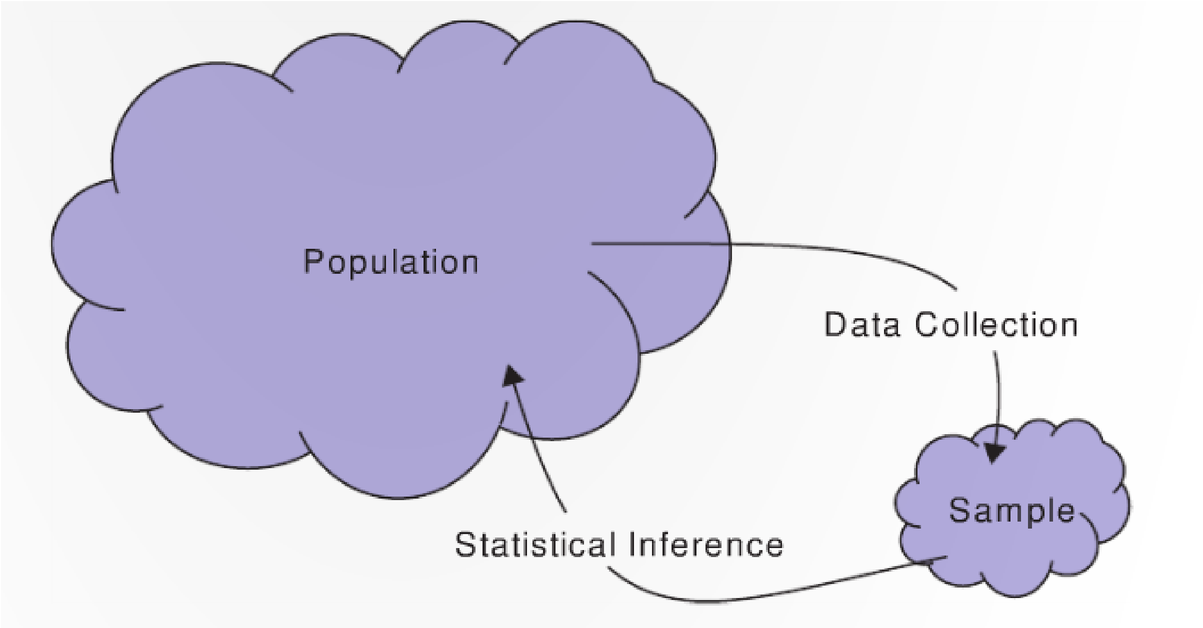


Sample: A subset of the population

Sampling

Simple random sample: each member in the population is equally likely to be in the sample

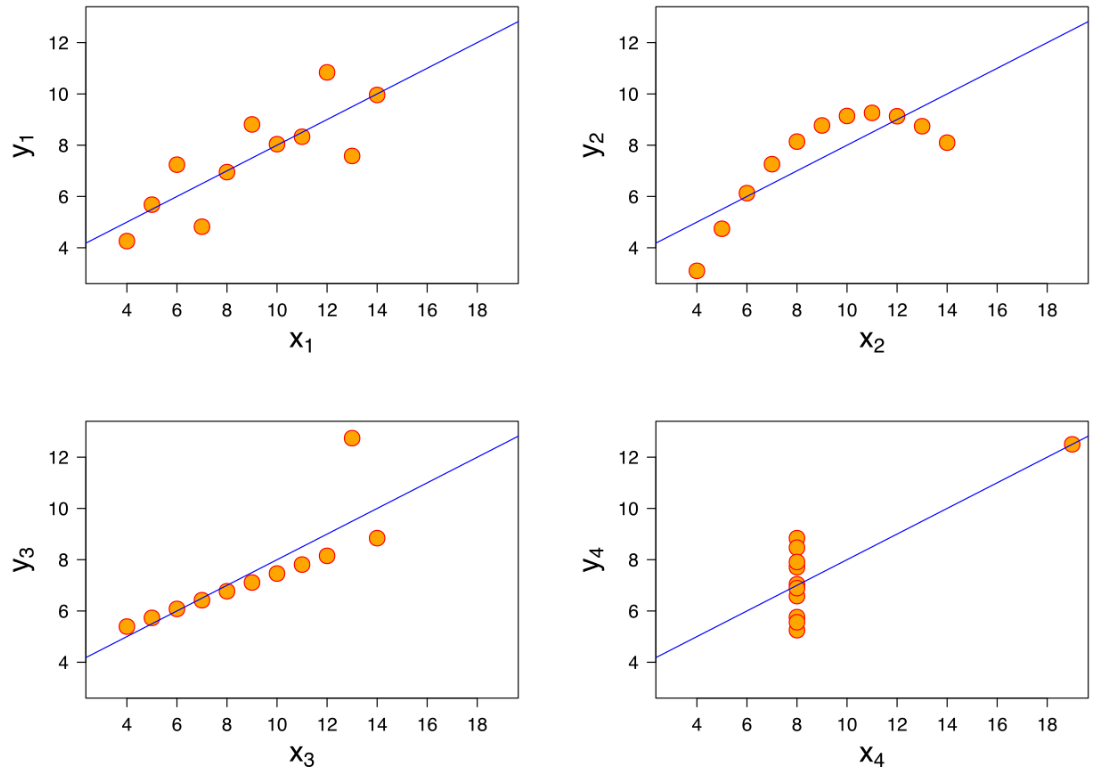
- This is called ***random selection***

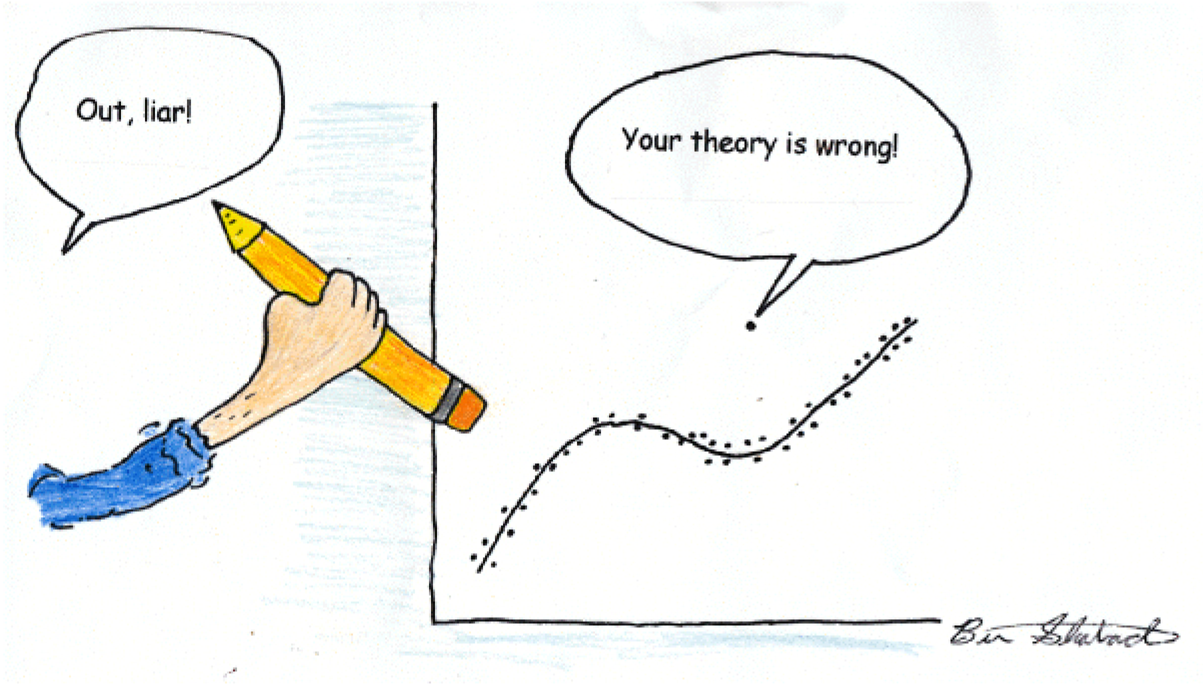


Soup analogy!

Why is it useful to use simple random sampling?

Outliers...





What do we do when we have outliers?

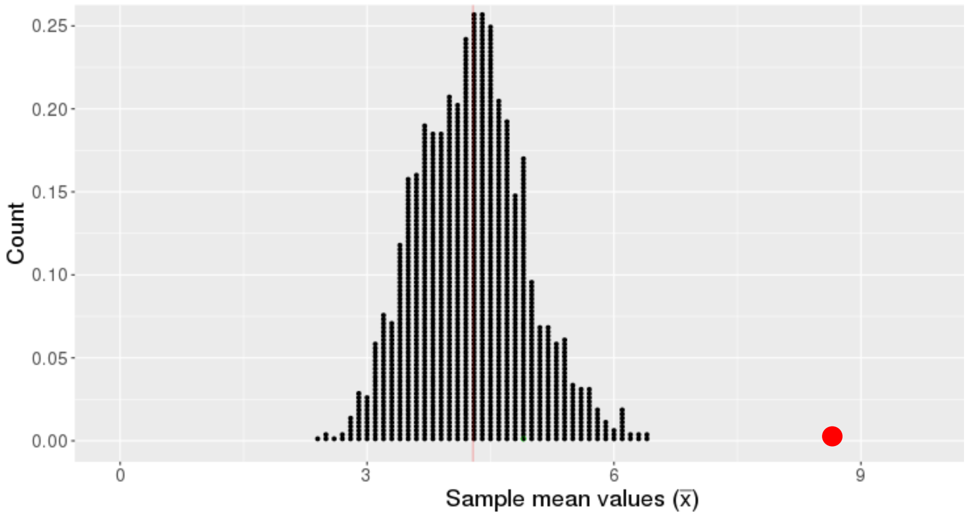
Basic hypothesis test logic

A **statistical test** uses data from a sample to assess a claim about a population

We start with a claim about a population parameter

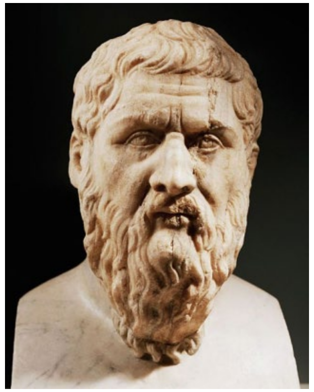
- E.g., $\mu = 4.2$

This claim implies we should get a certain distribution of statistics



If our observed statistic is highly unlikely, we reject the claim

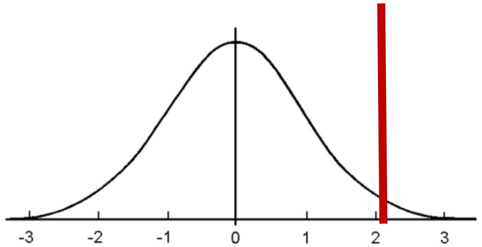
Five steps of hypothesis testing



1. State H_0 and H_A
   - Assume Gorgias (H_0) was right

2. Calculate the actual observed statistic

3. Create a distribution of what statistics would look like if Gorgias is right
   - Create the **null distribution** (that is consistent with H_0)

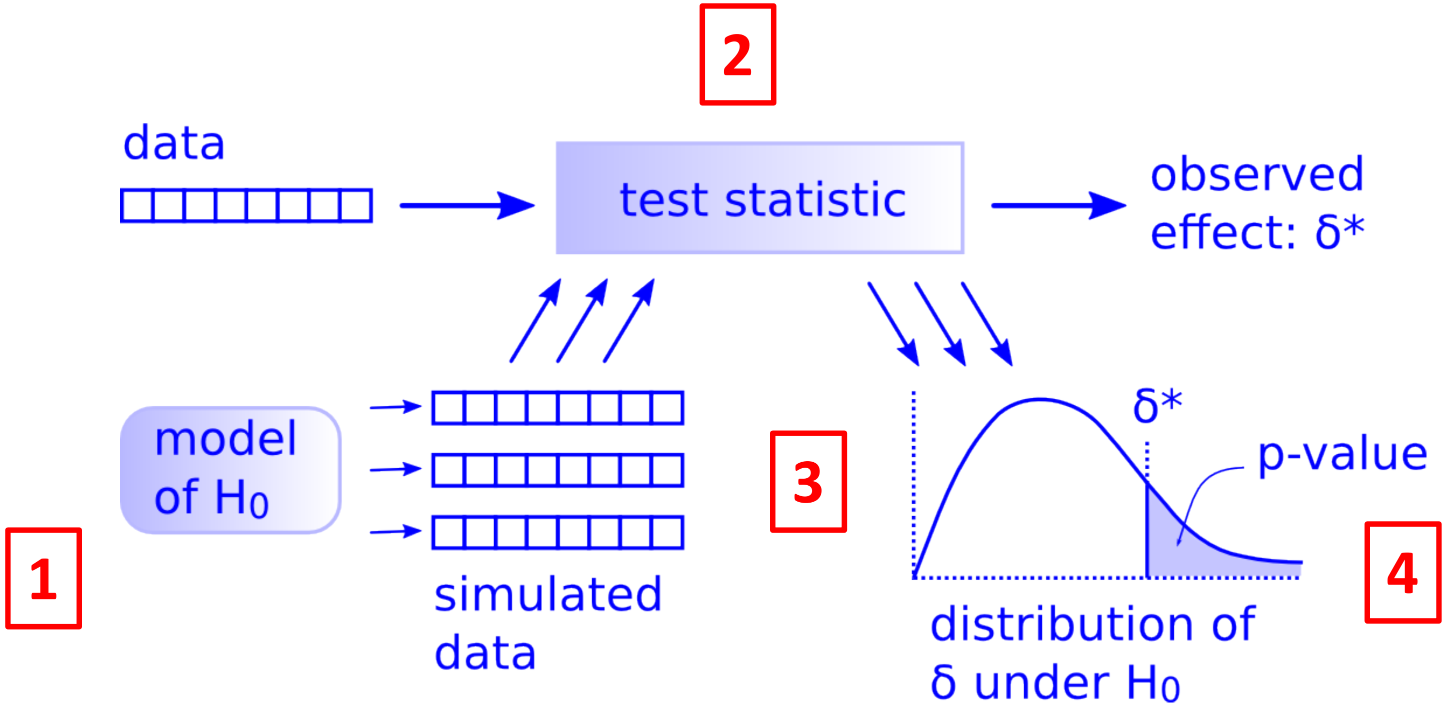


4. Get the probability we would get a statistic more than the observed statistic from the null distribution
   - p-value

5. Make a judgement
   - Assess whether the results are statistically significant

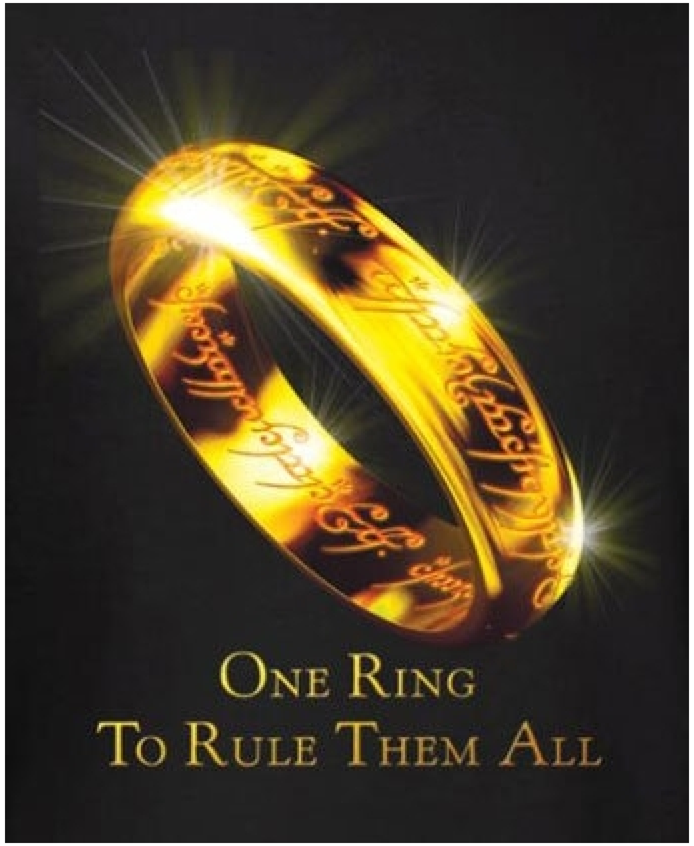
One test to rule them all

There is only one hypothesis test!



Just follow the 5 hypothesis tests steps!

5!



Five steps of hypothesis testing

1. State H_0 and H_A
   - Examples: H_0: $\pi = .5$ H_0: $\mu_1 - \mu_2 = 0$ H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

2. Calculate the actual observed statistic
   - Examples: $\hat{p} = .37$ $\bar{x}_1 - \bar{x}_2 = 2.27$ $F = 3.1$

3. Create a distribution of what statistics would look like if Gorgias is right
   - We can get a null distribution using:
     - **1. Randomization methods**
     - **2. Parametric density functions**

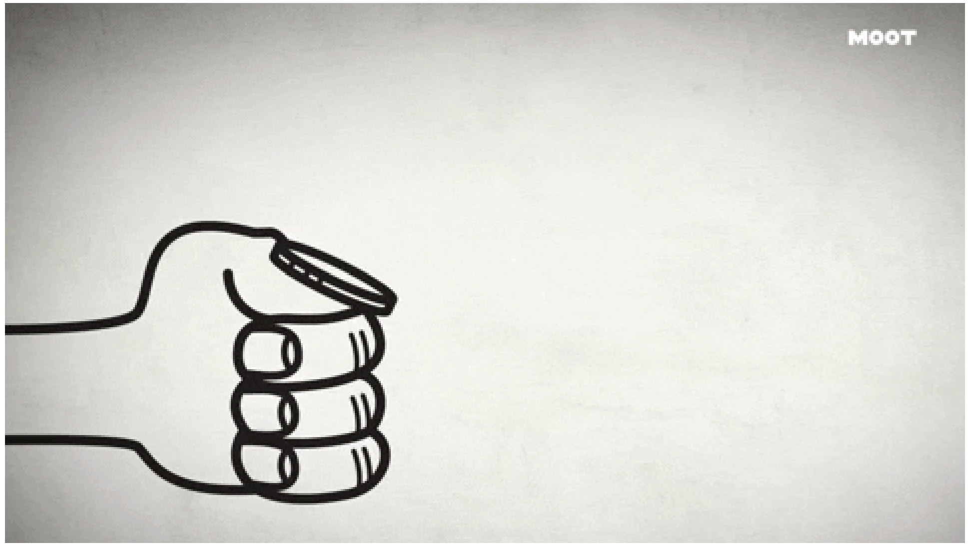
4. Get the probability we would get a statistic more than the observed statistic from the null distribution
   - p-value

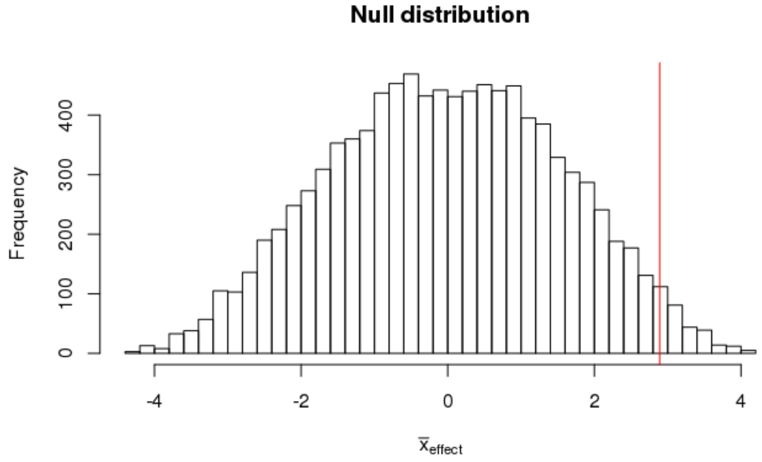
5. Make a judgement (if using the Neyman-Pearson paradigm)

Randomization null distributions (step 3)

For proportions, i.e., testing H_0: $\pi = \pi_0$

```
null_distribution <- rbinom(num_sims, size, prob)
```

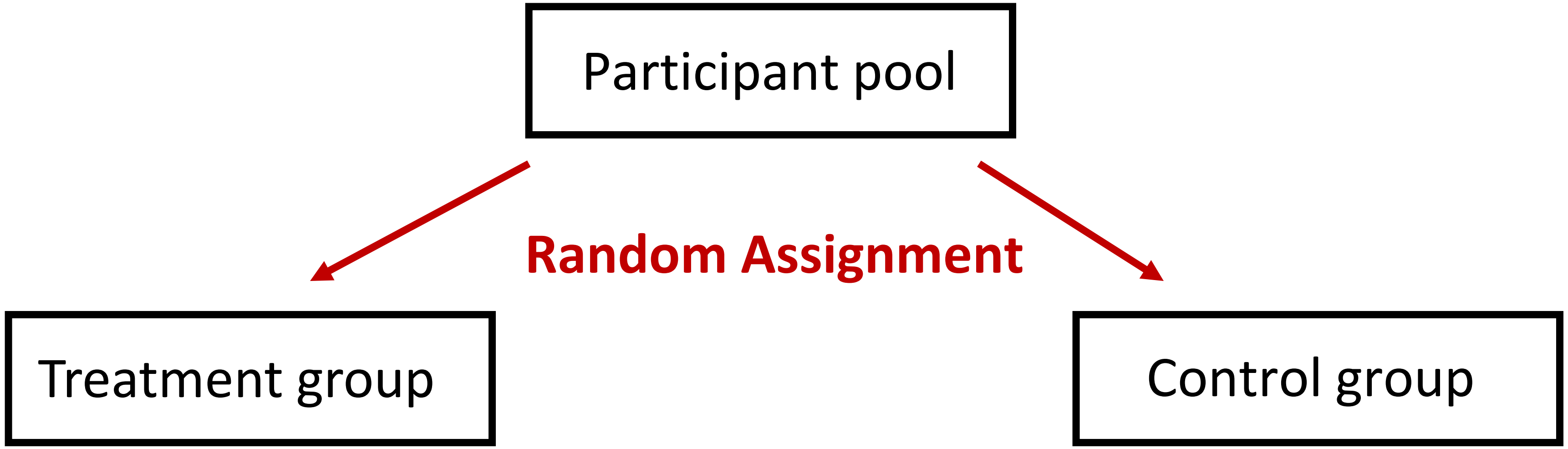




Experimental design

Take a group of participant and ***randomly assign***:

- Half to a *treatment group* where they get the pill
- Half in a *control group* where they get a fake pill (placebo)
- See if there is more improvement in the treatment group compared to the control group



3. Create the null distribution!

Treatment group

Control group

Reconstructed participant pool data under H_0

Shuffle data for random assignment consistent with H_0

Shuffled 'treatment group'

Shuffled 'control group'

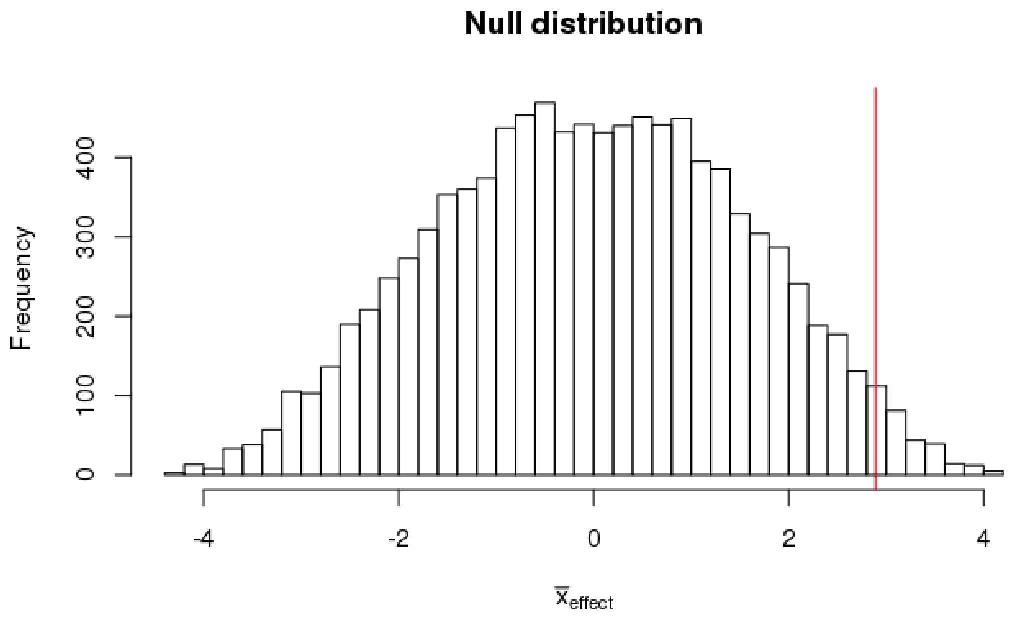
One null distribution statistic: $\bar{x}_{Shuff\_Treatment} - \bar{x}_{Shuff\_control}$

Randomization null distributions (step 3)

For comparing means, i.e., testing H_0: $\mu_1 - \mu_2 = 0$

```
combo_data <- c(group_data_1, group_data_2)
n1 <- length(group_1)
tot <- length(combo_data)
null_distribution <- NULL

for (i in 1:10000) {
        shuff_data <- sample(combo_data)
        shuff_group_1 <- shuff_data[1:n1]
        shuff_group_2 <- shuff_data[(n1 + 1):tot]
        null_distribution[i] <-
            mean(shuff_group_1) -
}
```



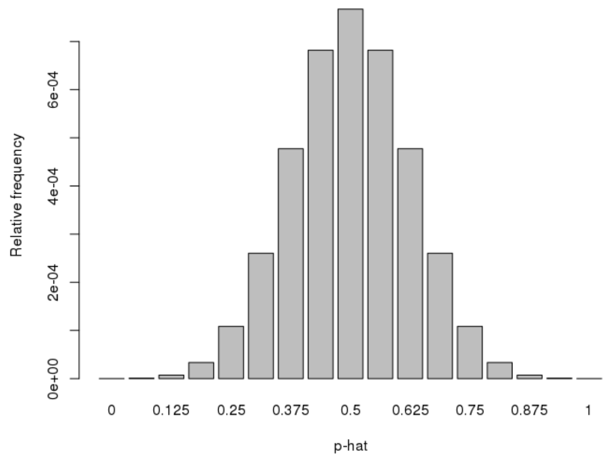

```
# one sided p-value
p_val <- c(null_distribution >= obs_stat)
```

Parametric null distributions (step 3)

For proportions, i.e., testing H_0: $\pi = \pi_0$

$$Pr(X = k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$$

```
null_distribution <- dbinom(x_range, size, prob)/size
```



$Pr(X \geq obs\_stat \mid H_0)$

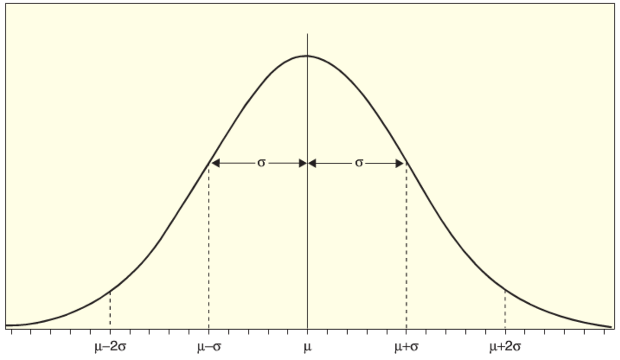
$Pr(X \geq obs\_stat \mid n, \pi = \pi_0)$

```
# one sided p-value
p_value <- pbinom(obs_num, size, prob)
```

Parametric null distributions (step 3)

For comparing means, i.e., testing H_0: $\mu_1 - \mu_2 = 0$

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$




```
y_vals <- dnorm(x_range, mu, sigma)
```

We usually use a t-statistic (t-test) for comparing means

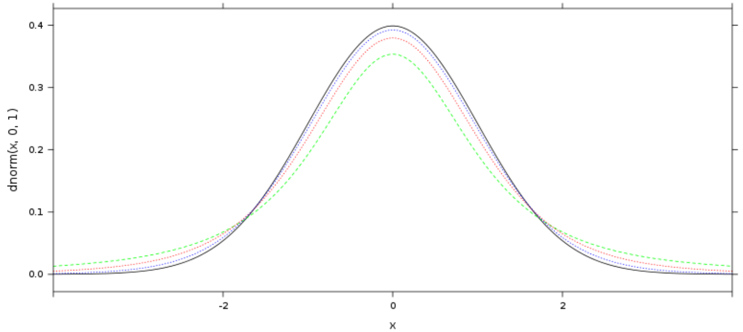
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\Pr(T \geq t\_stat \mid H_0)$

$\Pr(T \geq t\_stat \mid df)$

$df = \min(n_1, n_2) - 1$

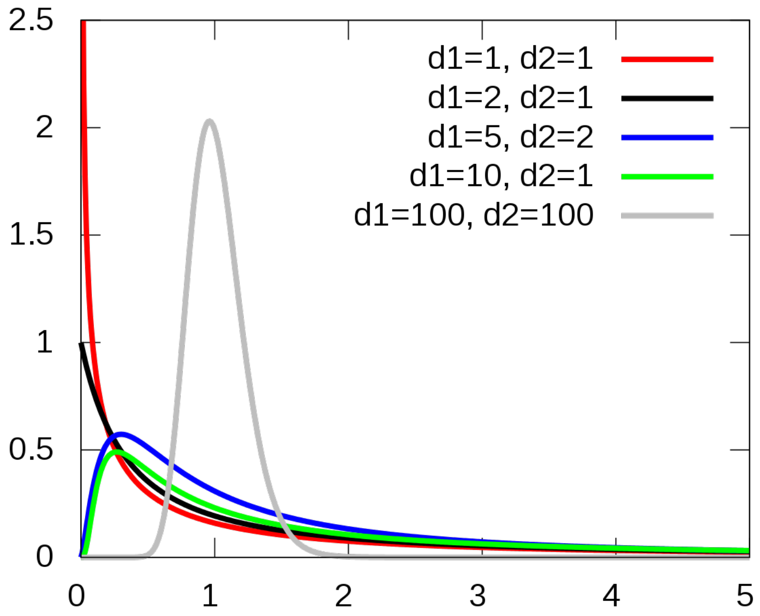
```
# one sided p-value
p_value <- pt(t_stat, df)
```



Parametric test for comparing more than one mean: One-way ANOVA

An Analysis of Variance (ANOVA) is a test that can be used to examine if a set of means are all the same

- H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
- $\mu_i \neq \mu_j$ for some i, j



The statistic we use for a one-way ANOVA is the F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

Any remaining questions about hypothesis tests?

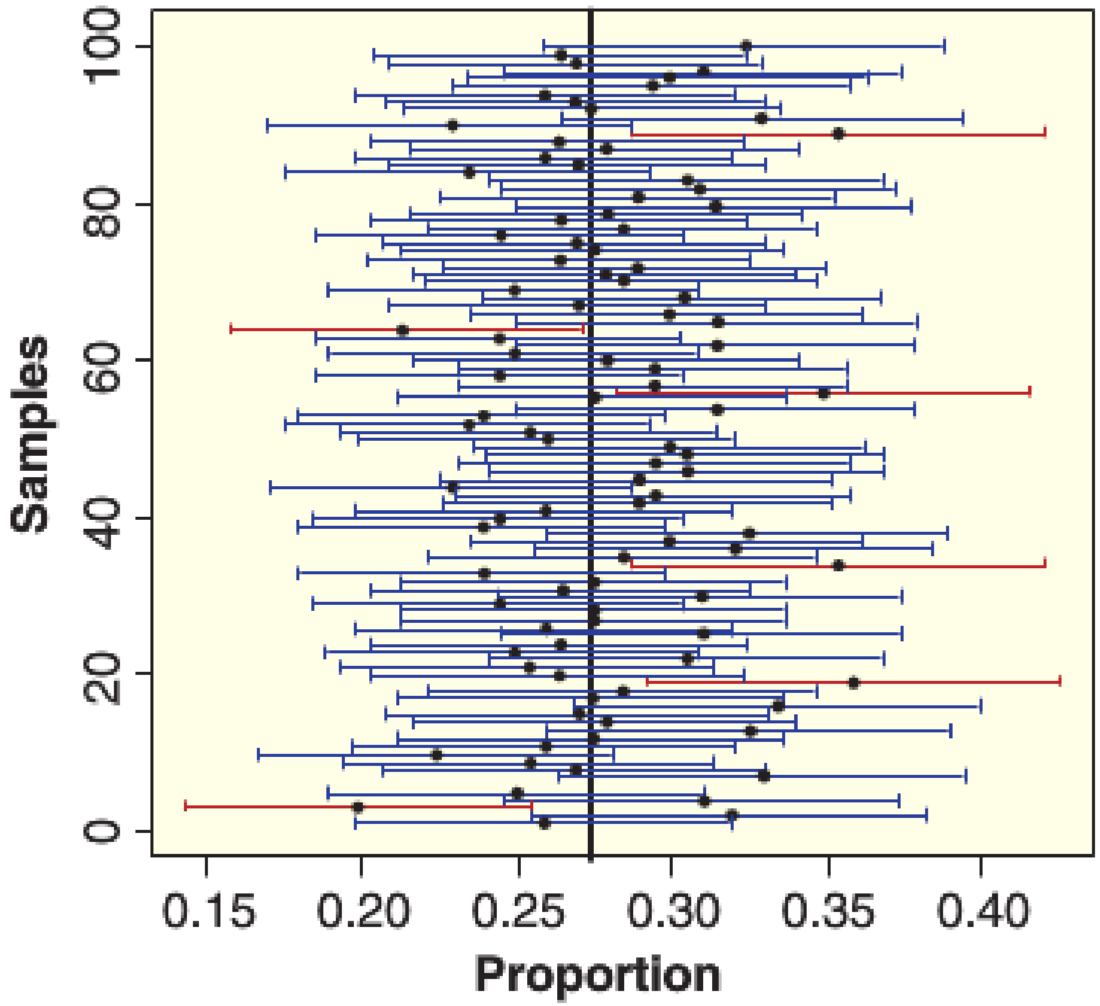
Confidence Intervals

For a **confidence level** of 95%...

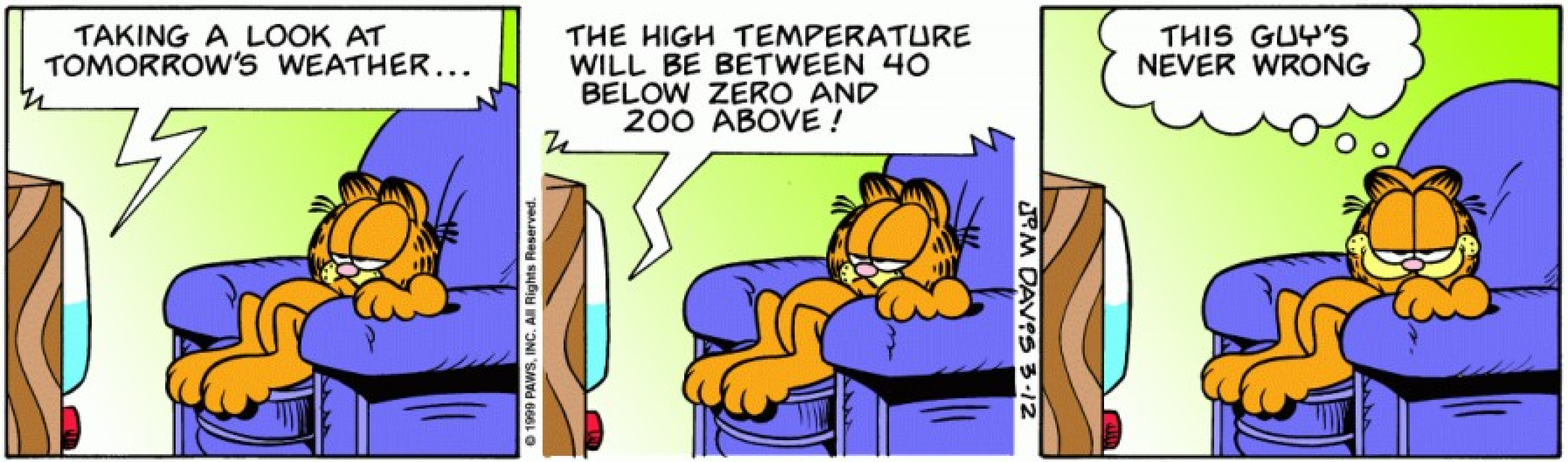
95% of the **confidence intervals** will have the ***parameter*** in them

Common form of a confidence interval:

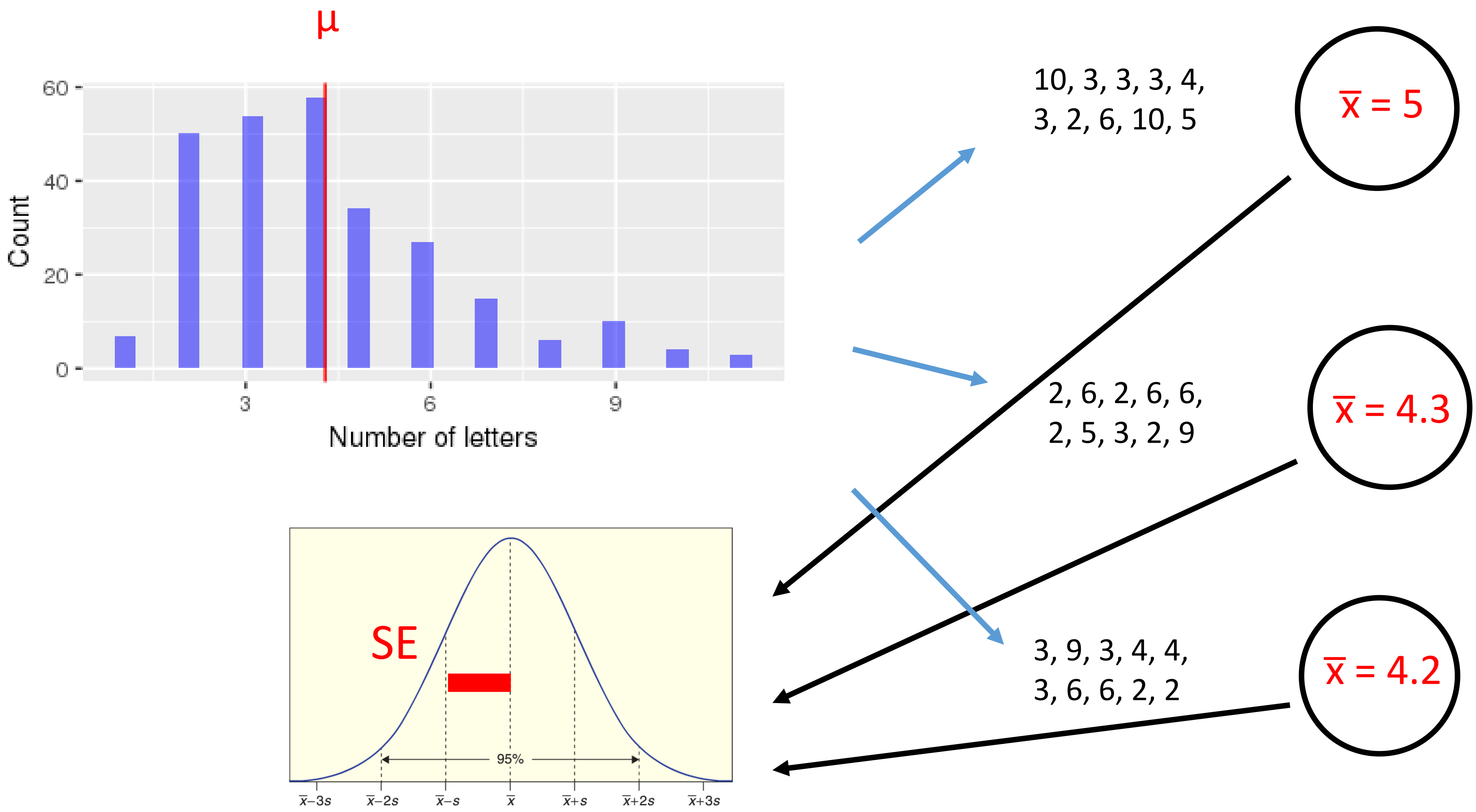
statistic ± q* $\cdot \hat{SE}$



There is a <u>tradeoff</u> between the **confidence level** (percent of times we capture the parameter) and the **confidence interval size**



Review: sampling distribution illustration

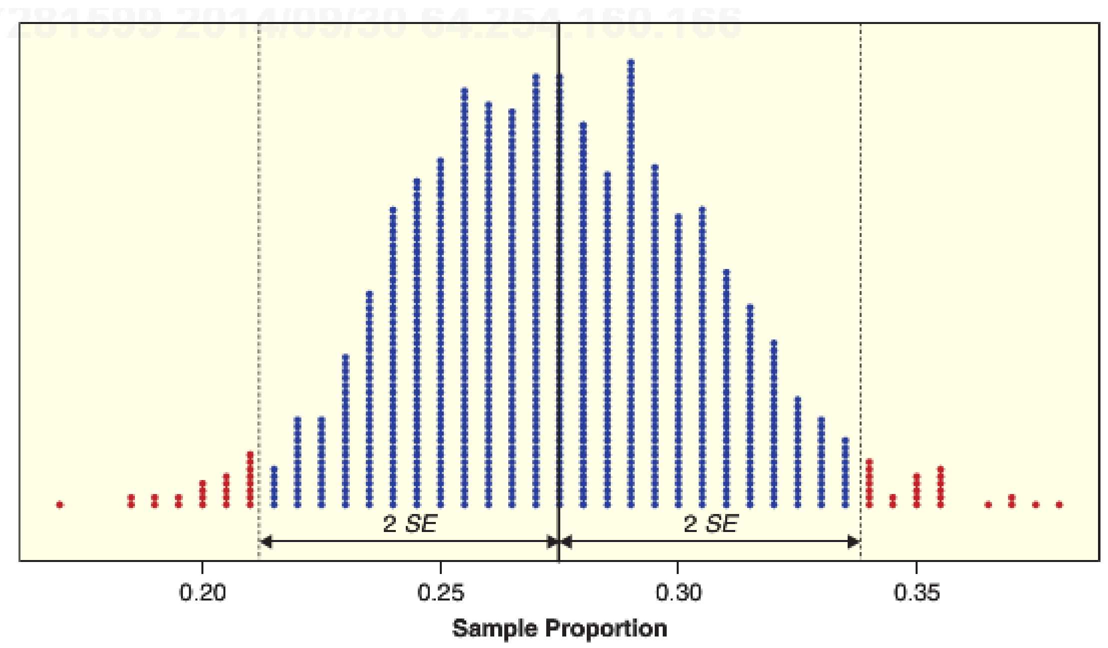


Sampling distribution!

A sampling distribution is a distribution of statistics

Sampling distributions

For a sampling distribution that is a normal distribution, 95% of ***statistics*** lie within 2 standard deviations (SE) for the population mean?

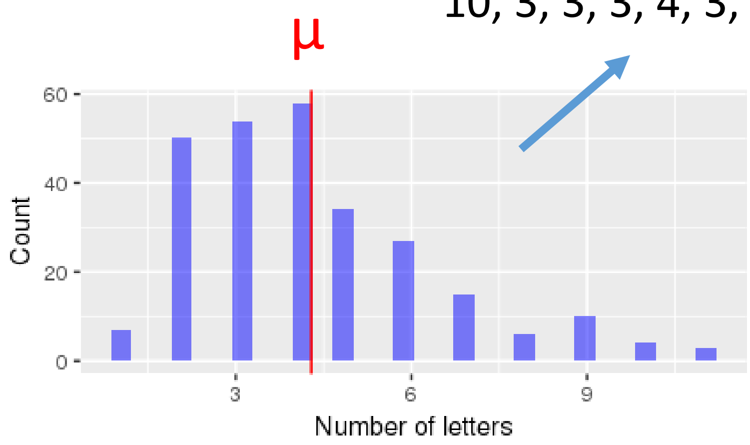


If we had a statistic value and the value of the SE could we compute a confidence interval using:

$$CI = stat \pm q^* \; SE$$

Bootstrap distribution

Sample with replacement!



<u>The sample (n = 10)</u>

10, 3, 3, 3, 4, 3, 2, 6, 4, 5

3, 3, 3, 5, 3,
4, 5, 2, 2, 10

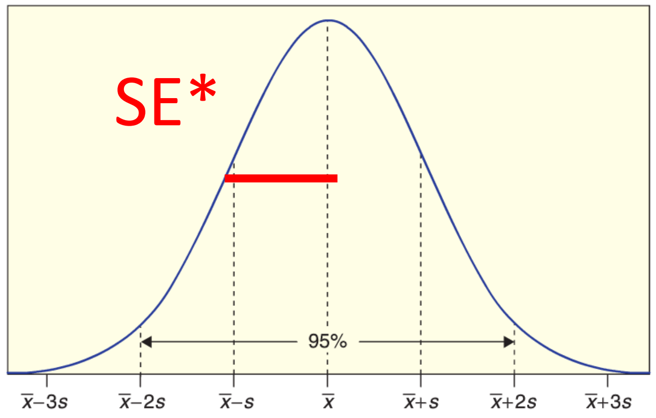
$\bar{x}^* = 4$

3, 3, 2, 3, 6,
4, 6, 5, 3, 6

$\bar{x}^* = 4.1$

5, 3, 2, 3, 3,
3, 10, 3, 4, 3

$\bar{x}^* = 3.9$



Bootstrap distribution!

Notice there is no 9's in the bootstrap samples

Bootstrap code

```
my_sample

bootstrap_dist <- NULL
 for (i in 1:10000) {
        boot_sample <- sample(my_sample, replace = TRUE)
        bootstrap_dist[i] <- mean(boot_sample)
}

boot_SE <- sd(bootstrap_dis)
```

Parametric formula for the standard error of the mean

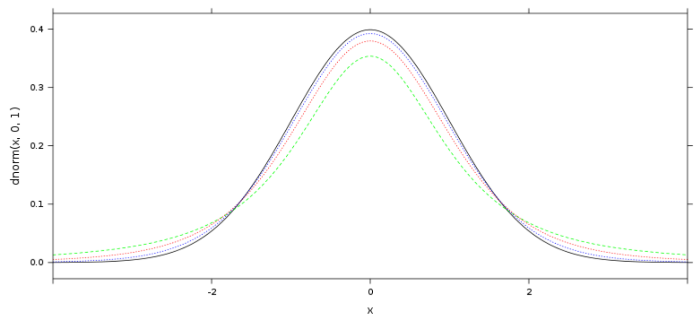
As you likely learned in intro statistics class, there is formula the **standard error of the mean (SEM)** which is:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad\qquad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Where:

- σ is population standard deviation parameter
- n is the sample size
- s is the sample standard deviation

sampling distribution



$\bar{x}$ (t)

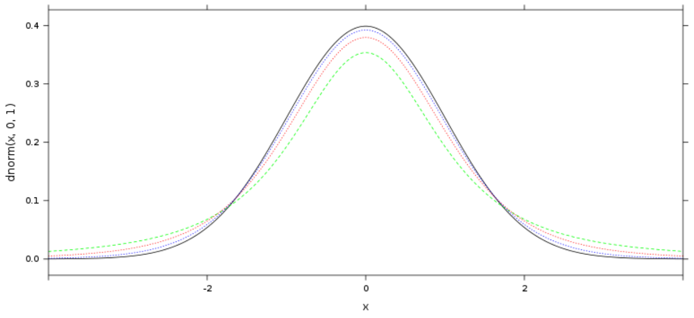
Confidence interval for μ: $\bar{x} \pm t^* \cdot s_{\bar{x}}$

Parametric formula for the standard error of a proportion

Likewise, there is a formula for **standard error of a proportion** which is:

$$\sigma_{\hat{p}} = \sqrt{\frac{\pi \cdot (1-\pi)}{n}} \qquad s_{\hat{p}} = \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

sampling distribution



$\hat{p}$

Where:

- $\hat{\pi}$ is the population proportion parameter
- n is the sample size
- $\hat{p}$ is the sample proportion statistic

Two theories of hypothesis testing

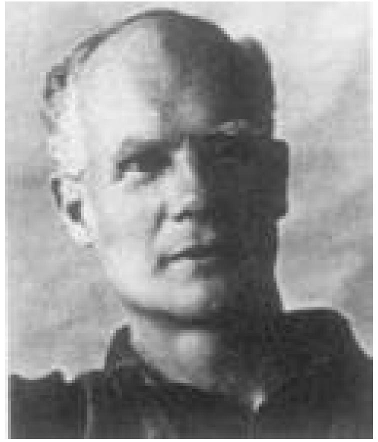
Null-hypothesis significance testing (NHST) is a hybrid of two theories:

1. Significance testing of Ronald Fisher

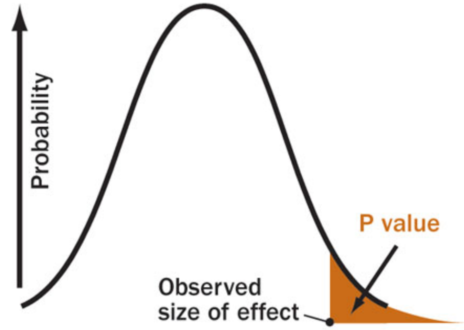
2. Hypothesis testing of Jezy Neyman and Egon Pearson

Fisher (1890-1962)

Neyman (1894-1981)

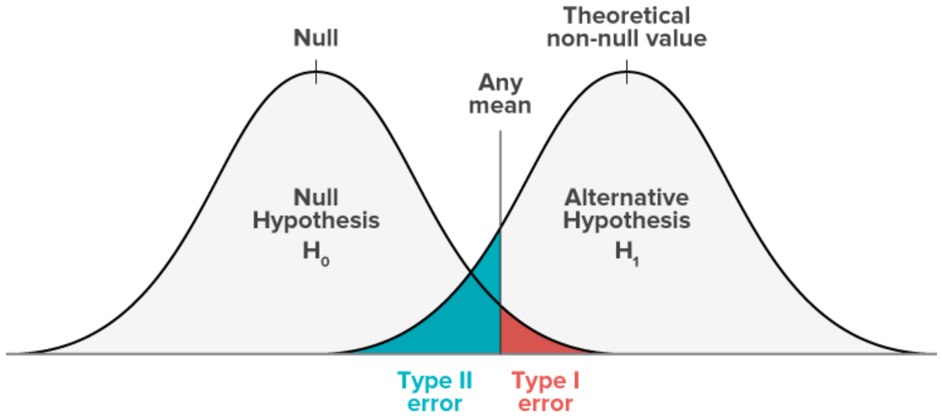


Pearson (1895-1980)



Problems with the NP hypothesis tests

<u>Problem 1</u>: we are interested in the results of a specific experiment, not whether we are right most of the time



<u>Problem 2</u>: Arbitrary thresholds for alpha levels

<u>Problem 3</u>: running many tests can give rise to a high number of type 1 errors