INMAS

Hypothesis Testing

- *Testing Frameworks*
- *Ides of Testing*

James Balamuta

INMAS Statistical Methods Workshop Fall 2021

Lecture Objectives

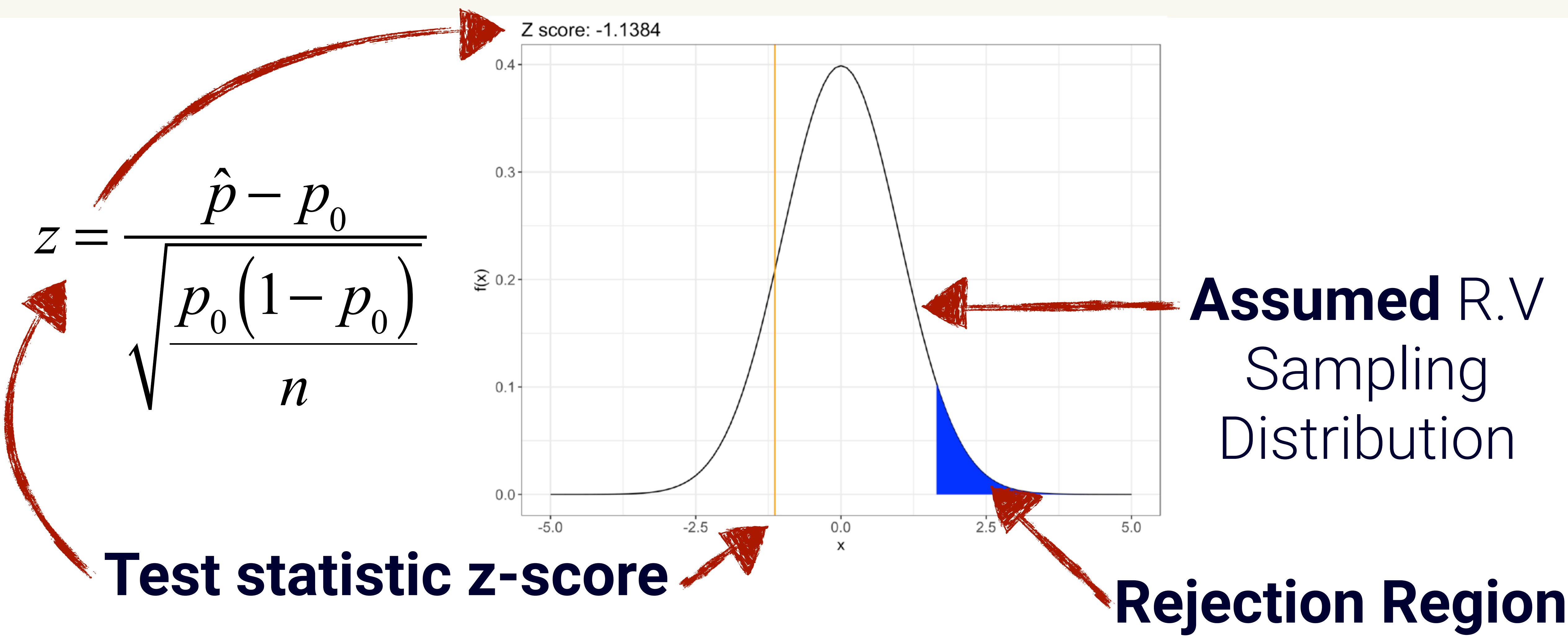
- **Describe** how hypothesis testing is part of confirmatory data analysis and the importance of inference.
- **Differentiate** between different hypothesis testing frameworks.

Testing Frameworks

Definition:

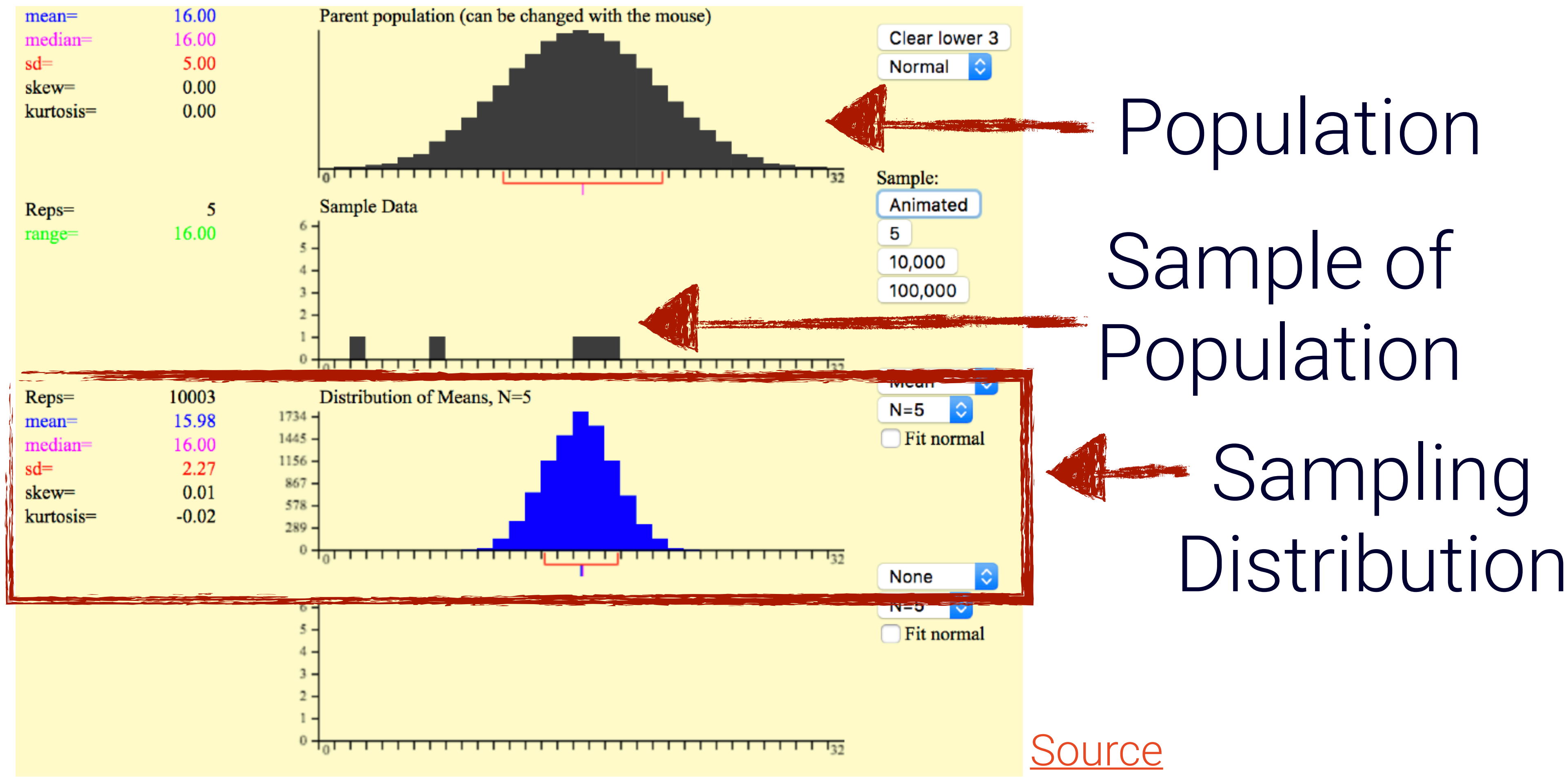
Hypothesis testing or *confirmatory data analysis* is the act of examining whether random variables (RVs) adhere to stated assumptions or differ.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$



Definition:

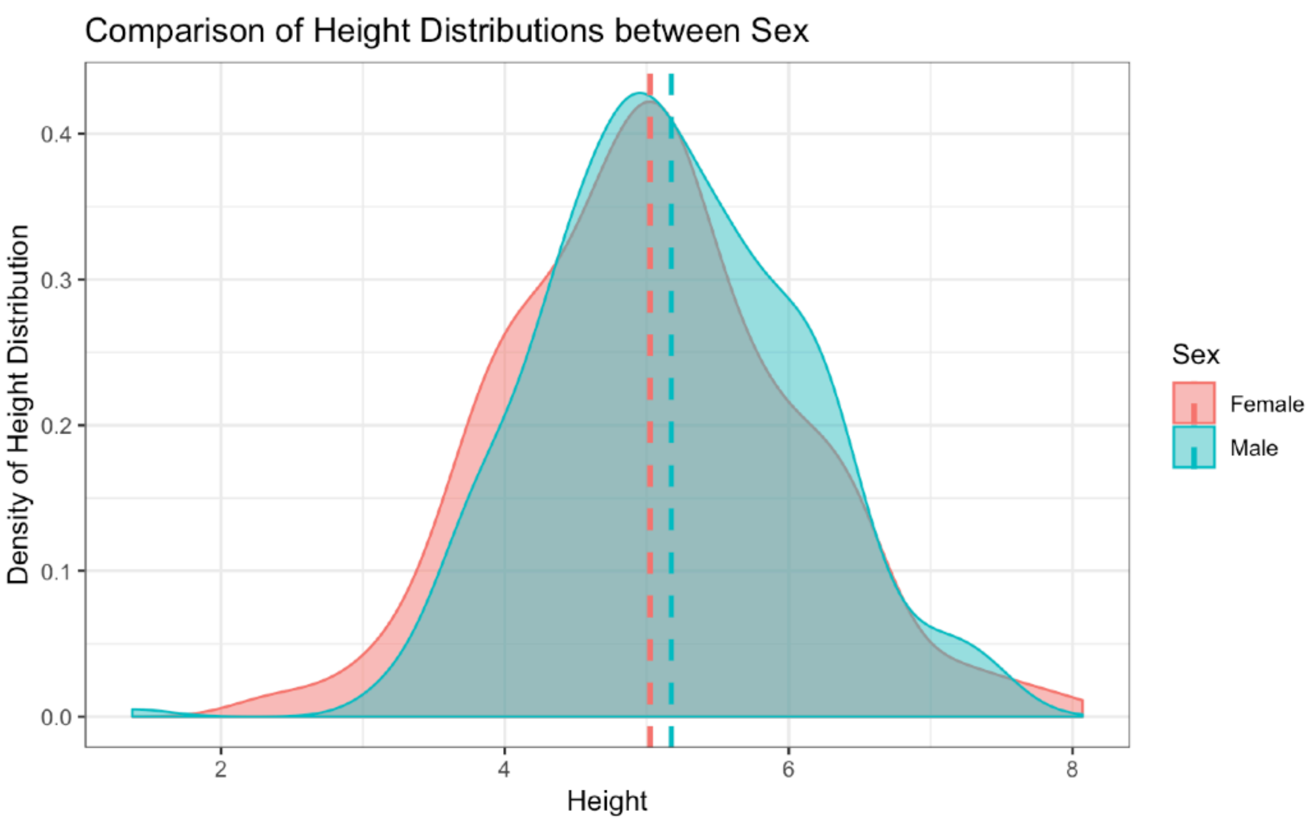
Sampling Distribution is a probability distribution of statistics obtained from drawing many samples from a population of interest.

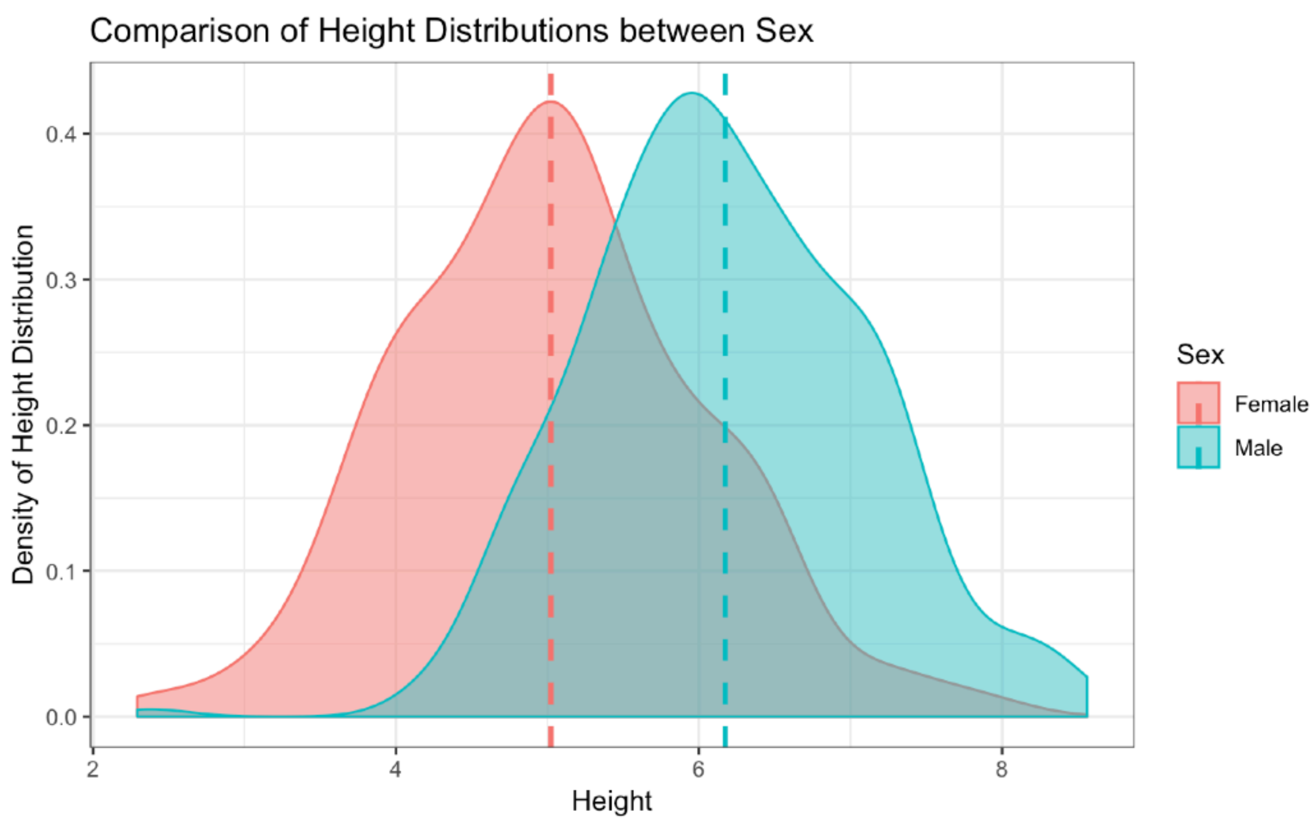


Source

Usefulness of Tests

… does one group differ from the next ???





vs.

Computation Types

… how can we test our assumptions ??

- **Parametric**
  - Assume a distribution and compute a test statistic based on asymptotic results.
- **Resampling**
  - Draw samples **with** replacement from the sampling distribution.
- **Permutation**
  - Shuffle the samples and sample **without** replacement.

Ides of Testing

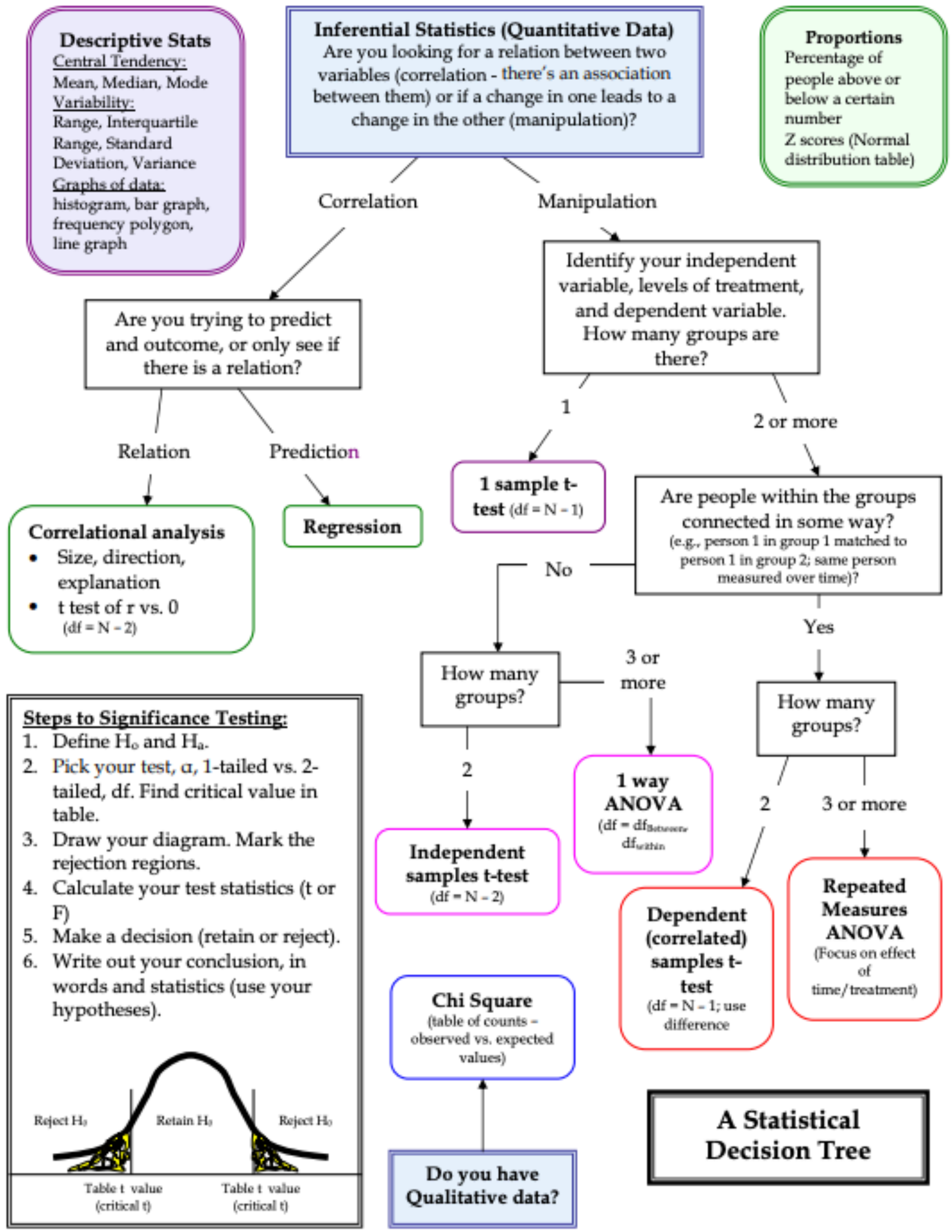
Hypothesis Tests

Common Parametric Hypothesis Tests

Hypotheses	Assumptions	Critical Region
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$N(\mu, \sigma^2)$ or n large, σ^2 known	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$N(\mu, \sigma^2)$ σ^2 unknown	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq t_\alpha(n-1)$
$H_0: \mu_X - \mu_Y = 0$ $H_1: \mu_X - \mu_Y > 0$	$N(\mu_X, \sigma_X^2)$ $N(\mu_Y, \sigma_Y^2)$ σ_X^2, σ_Y^2 known	$z = \dfrac{\bar{x} - \bar{y} - 0}{\sqrt{(\sigma_X^2/n) + (\sigma_Y^2/m)}} \geq z_\alpha$
$H_0: \mu_X - \mu_Y = 0$ $H_1: \mu_X - \mu_Y > 0$	Variances unknown, large samples	$z = \dfrac{\bar{x} - \bar{y} - 0}{\sqrt{(s_x^2/n) + (s_y^2/m)}} \geq z_\alpha$
$H_0: \mu_X - \mu_Y = 0$ $H_1: \mu_X - \mu_Y > 0$	$N(\mu_X, \sigma_X^2)$ $N(\mu_Y, \sigma_Y^2)$ $\sigma_X^2 = \sigma_Y^2$, unknown	$t = \dfrac{\bar{x} - \bar{y} - 0}{s_p\sqrt{(1/n) + (1/m)}} \geq t_\alpha(n+m-2)$ $s_p = \sqrt{\dfrac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}$
$H_0: \mu_D = \mu_X - \mu_Y = 0$ $H_1: \mu_D = \mu_X - \mu_Y > 0$	X and Y normal, but dependent	$t = \dfrac{\bar{d} - 0}{s_d/\sqrt{n}} \geq t_\alpha(n-1)$
$H_0: p = p_0$ $H_1: p > p_0$	$b(n, p)$ n is large	$z = \dfrac{(y/n) - p_0}{\sqrt{p_0(1-p_0)/n}} \geq z_\alpha$
$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 > 0$	$b(n_1, p_1)$ $b(n_2, p_2)$	$z = \dfrac{(y_1/n_1) - (y_2/n_2) - 0}{\sqrt{\left(\dfrac{y_1+y_2}{n_1+n_2}\right)\left(1 - \dfrac{y_1+y_2}{n_1+n_2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \geq z_\alpha$

Back cover of: Probability and Statistical Inference (9th Edition)

Decision Hierarchy



Source

Common statistical tests are linear models

Last updated: 28 June, 2019.Also check out the Python version!

See worked examples and more details at the accompanying notebook: https://lindeloev.github.io/tests-as-linear

	Common name	Built-in function in R	Equivalent linear model in R	Exact?	The linear model in words	Icon
Simple regression: lm(y ~ 1 + x)	**y is independent of x** P: One-sample t-test N: Wilcoxon signed-rank	t.test(y) wilcox.test(y)	lm(y ~ 1) lm(signed_rank(y) ~ 1)	✓ for N >14	One number (intercept, i.e., the mean) predicts **y**. - (Same, but it predicts the *signed rank* of **y**.)	
	P: Paired-sample t-test N: Wilcoxon matched pairs	t.test(y_1, y_2, paired=TRUE) wilcox.test(y_1, y_2, paired=TRUE)	lm(y_2 - y_1 ~ 1) lm(signed_rank(y_2 - y_1) ~ 1)	✓ for N >14	One intercept predicts the pairwise **y_2-y_1** differences. - (Same, but it predicts the *signed rank* of **y_2-y_1**.)	
	y ~ continuous x P: Pearson correlation N: Spearman correlation	cor.test(x, y, method='Pearson') cor.test(x, y, method='Spearman')	lm(y ~ 1 + x) lm(rank(y) ~ 1 + rank(x))	✓ for N >10	One intercept plus **x** multiplied by a number (slope) predicts **y**. - (Same, but with *ranked* **x** and **y**)	
	y ~ discrete x P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	t.test(y_1, y_2, var.equal=TRUE) t.test(y_1, y_2, var.equal=FALSE) wilcox.test(y_1, y_2)	lm(y ~ 1 + G_2)[A] gls(y ~ 1 + G_2, weights=...[B])[A] lm(signed_rank(y) ~ 1 + G_2)[A]	✓ ✓ for N >11	An intercept for **group 1** (plus a difference if **group 2**) predicts **y**. - (Same, but with one variance *per group* instead of one common.) - (Same, but it predicts the *signed rank* of **y**.)	
Multiple regression: lm(y ~ 1 + x_1 + x_2 +...)	P: One-way ANOVA N: Kruskal-Wallis	aov(y ~ group) kruskal.test(y ~ group)	lm(y ~ 1 + G_2 + G_3 +...+ G_N)[A] lm(rank(y) ~ 1 + G_2 + G_3 +...+ G_N)[A]	✓ for N >11	An intercept for **group 1** (plus a difference if group ≠ 1) predicts **y**. - (Same, but it predicts the *rank* of **y**.)	
	P: One-way ANCOVA	aov(y ~ group + x)	lm(y ~ 1 + G_2 + G_3 +...+ G_N + x)[A]	✓	- (Same, but plus a slope on **x**.) *Note: this is discrete AND continuous. ANCOVAs are ANOVAs with a continuous x.*	
	P: Two-way ANOVA	aov(y ~ group * sex)	lm(y ~ 1 + G_2 + G_3 +...+ G_N + S_2 + S_3 +...+ S_K + G_2*S_2 + G_3*S_3 + ... + G_N*S_K)	✓	Interaction term: changing **sex** changes the **y ~ group** parameters. *Note: **$G_{2 to N}$** is an indicator (0 or 1) for each non-intercept levels of the **group** variable. Similarly for **$S_{2 to K}$** for sex. The first line (with G_i) is main effect of group, the second (with S_i) for sex and the third is the **group × sex** interaction. For two levels (e.g. male/female), line 2 would just be "S_2" and line 3 would be S_2 multiplied with each G_i.*	*[Coming]*
	Counts ~ discrete x N: Chi-square test	chisq.test(groupXsex_table)	**Equivalent log-linear model** glm(y ~ 1 + G_2 + G_3 + ... + G_N + S_2 + S_3+ ... + S_K + G_2*S_2 + G_3*S_3 +...+ G_N*S_K, family=...)[A]	✓	Interaction term: (Same as Two-way ANOVA.) *Note: Run glm using the following arguments: `glm(model, family=poisson())` As linear-model, the Chi-square test is $\log(y_i) = \log(N) + \log(\alpha_i) + \log(\beta_j) + \log(\alpha_i\beta_j)$ where α_i and β_j are proportions. See more info in the accompanying notebook.*	*Same as Two-way ANOVA*
	N: Goodness of fit	chisq.test(y)	glm(y ~ 1 + G_2 + G_3 +...+ G_N, family=...)[A]	✓	(Same as One-way ANOVA and see Chi-Square note.)	*1W-ANOVA*

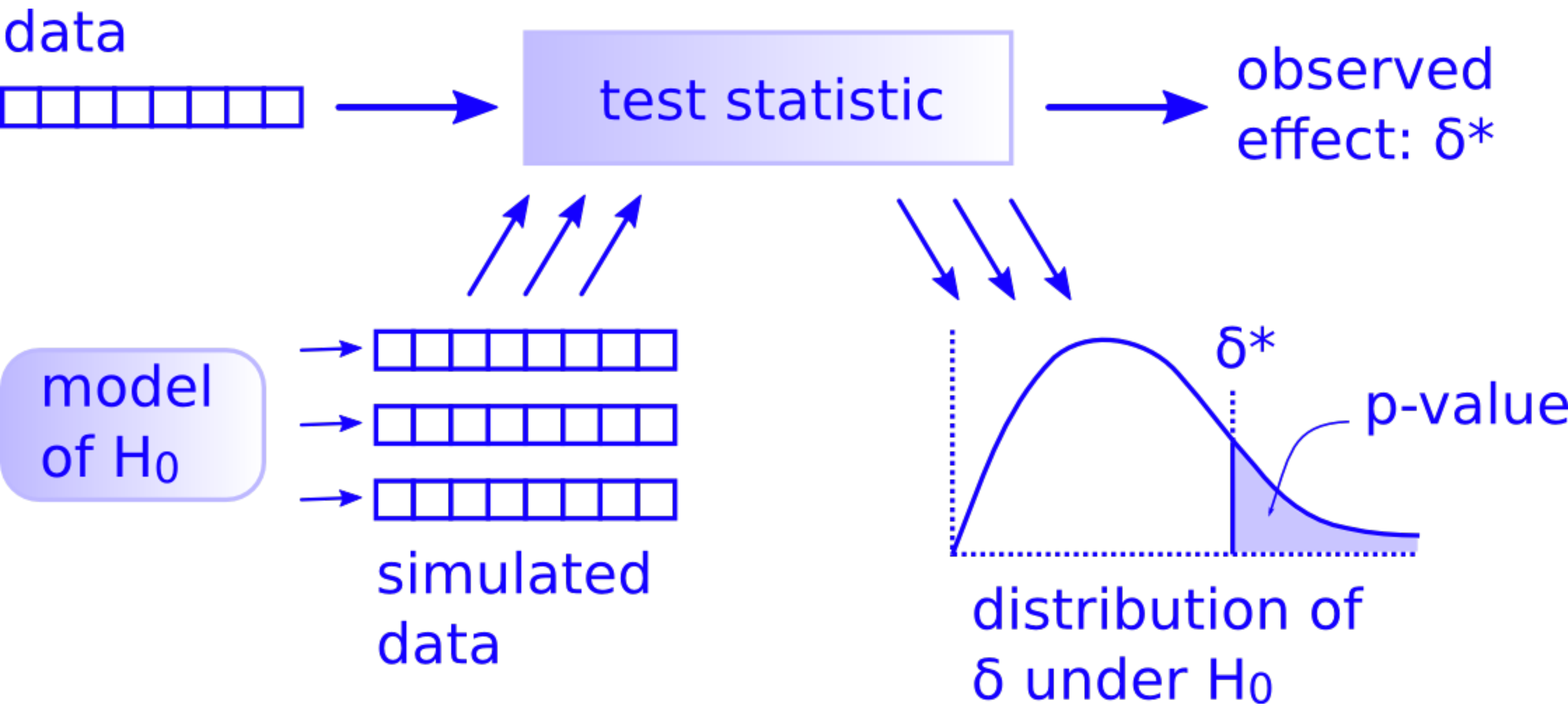
List of common parametric (P) non-parametric (N) tests and equivalent linear models. The notation y ~ 1 + x is R shorthand for $y = 1 \cdot b + a \cdot x$ which most of us learned in school. Models in similar colors are highly similar, but really, notice how similar they *all* are across colors! For non-parametric models, the linear models are reasonable approximations for non-small sample sizes (see "Exact" column and click links to see simulations). Other less accurate approximations exist, e.g., Wilcoxon for the sign test and Goodness-of-fit for the binomial test. The signed rank function is `signed_rank = function(x) sign(x) * rank(abs(x))`. The variables G_i and S_i are "dummy coded" indicator variables (either 0 or 1) exploiting the fact that when $\Delta x = 1$ between categories the difference equals the slope. Subscripts (e.g., G_2 or y_1) indicate different columns in data. lm requires long-format data for all non-continuous models. All of this is exposed in greater detail and worked examples at https://lindeloev.github.io/tests-as-linear.

[A] See the note to the two-way ANOVA for explanation of the notation.
[B] Same model, but with one variance per group: `gls(value ~ 1 + G2, weights = varIdent(form = ~1|group), method="ML")`.

Jonas Kristoffer Lindeløv
https://lindeloev.net

https://lindeloev.github.io/tests-as-linear/

"There is Only One Test"



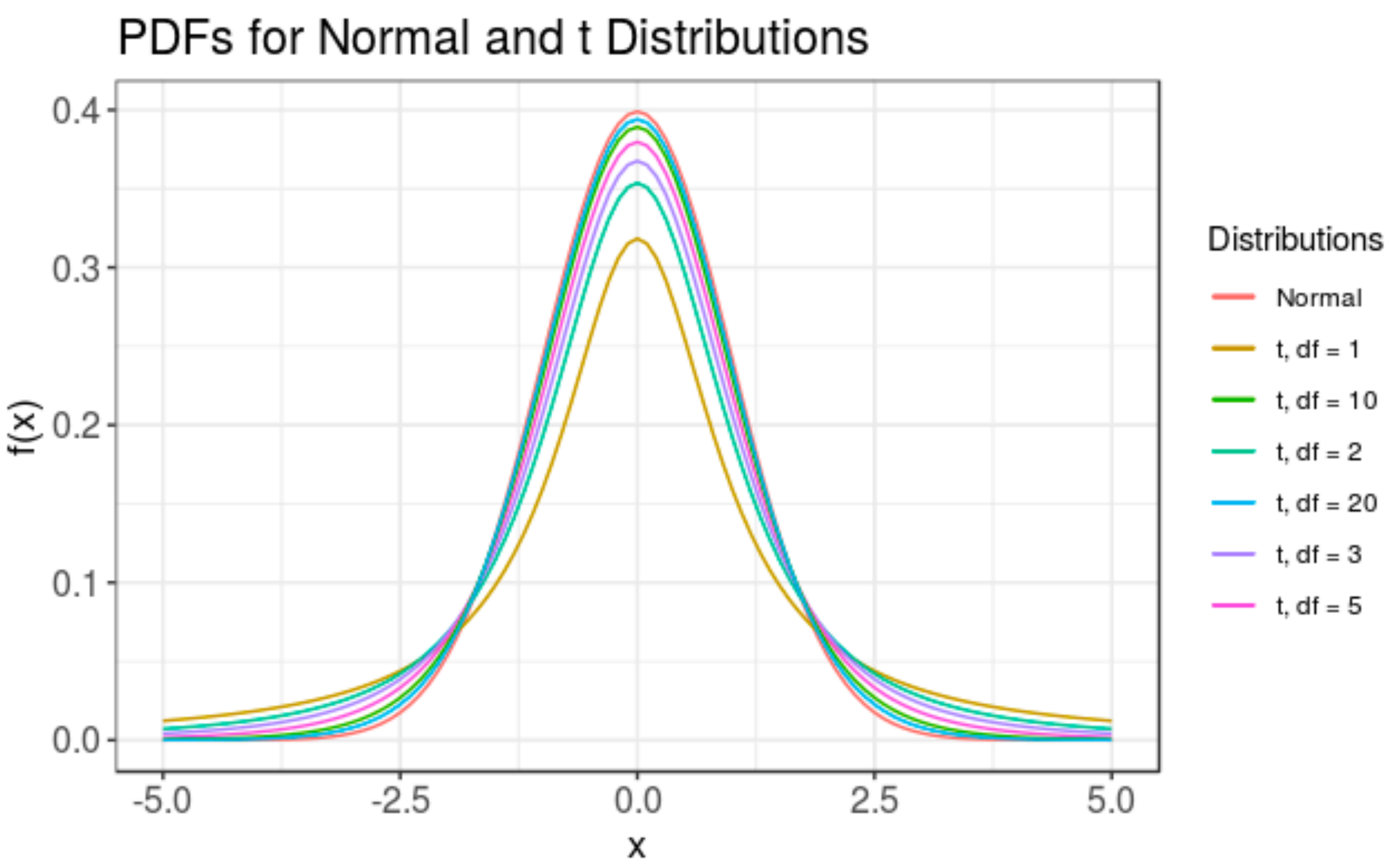
http://allendowney.blogspot.com/2011/05/there-is-only-one-test.html

Breaking Down All Tests

… the logic of a hypothesis test …

1. **Make an assumption or null hypothesis…**
   … about the sampling distribution the data has …
2. **Compute a Test Statistic …**
   … Relevant to that sampling distribution …
3. **Compute a *P*-value or Critical Value …**
   … probability of an effect as big given the distribution …
4. **Make a decision …**
   … does evidence exist to suggest the assumption is not okay?

Make an assumption…



Choose between:

- **z-test**
  - *Normal distribution*
  - Known Population Variance
  - Sample size: n > 30
- **t-test**
  - *Student's t*
  - Unknown Population Variance
  - Sample size: n < 30

Definition:
Null Hypothesis (H_0) states that the sample data gathered meets the sampling distributions assumptions.

	Means	**Proportions**
Single	$H_0 : \mu = \mu_0$	$H_0 : p = p_0$
Double	$H_0 : \mu_1 - \mu_2 = 0$	$H_0 : p_1 - p_2 = 0$

No difference

Definition:
Alternative Hypothesis (H_a or H_1) is the counter to the null hypothesis that emphasizes the underlying sample *may not* follow the assumptions laid out.

	Means	Proportions
One-sided	$H_a: \mu_1 - \mu_2 > 0$ $H_a: \mu > \mu_0$	$H_a: p > p_0$ $H_a: p < p_0$
Two-sided	$H_a: \mu_1 - \mu_2 \neq 0$ $H_a: \mu \neq \mu_0$	$H_a: p_1 - p_2 \neq 0$ $H_a: p \neq p_0$

Evidence suggests a difference

Example Hypothesis

… sample of forming a statement …

- Consider a coin flip.
- If the coin is fair, half the flips should be heads and the other half tails
- Otherwise, the number of flips between heads and tails will differ.

"Heads" "Tails"

$$H_0 : p = 0.5$$

$$H_a : p \neq 0.5$$

One proportion z-test

Addressing a probability

Null —• $H_0 : p = 0.5$

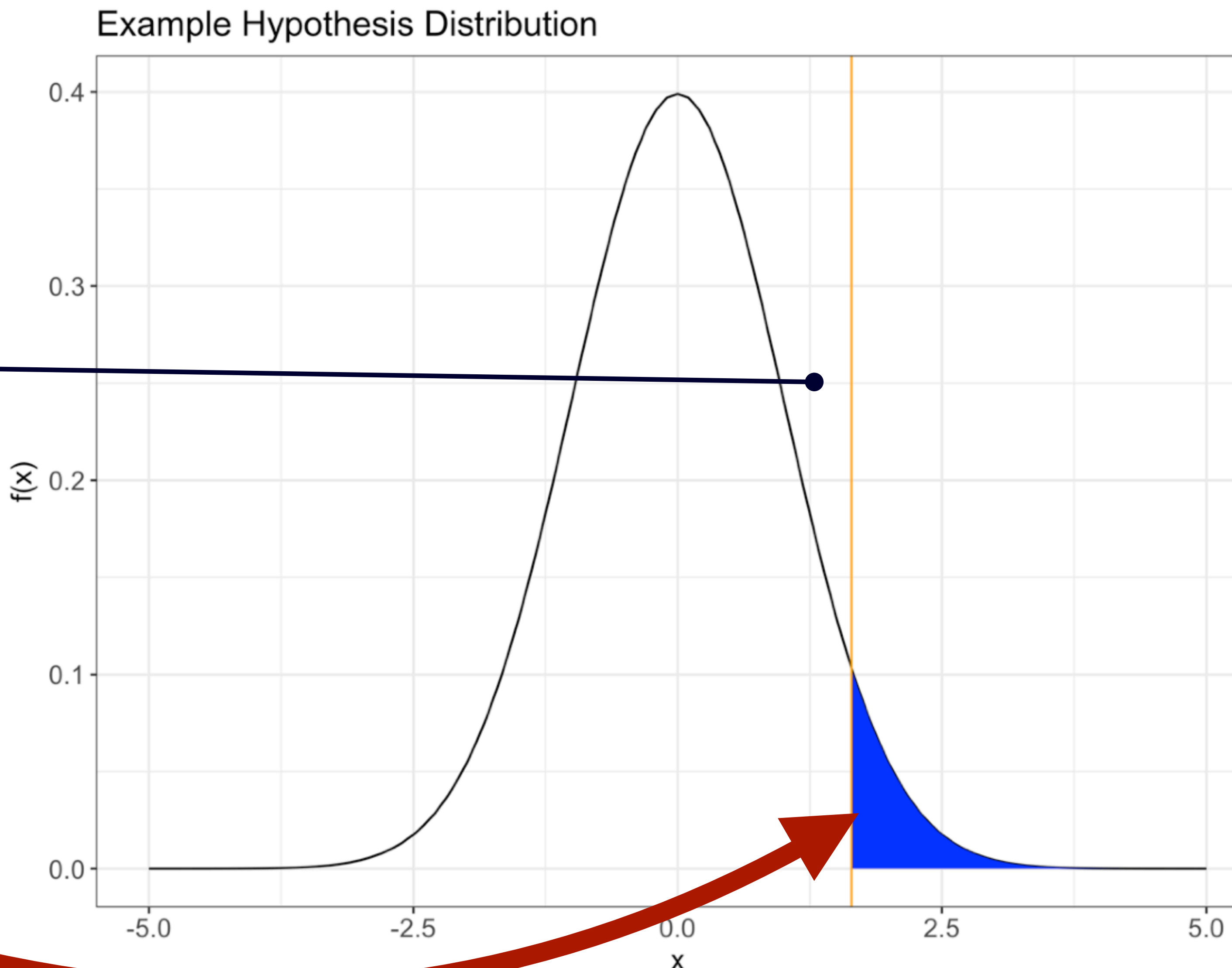
$H_a : p > 0.5$ •— Alternative

$$\hat{p} = \frac{observed}{total}$$

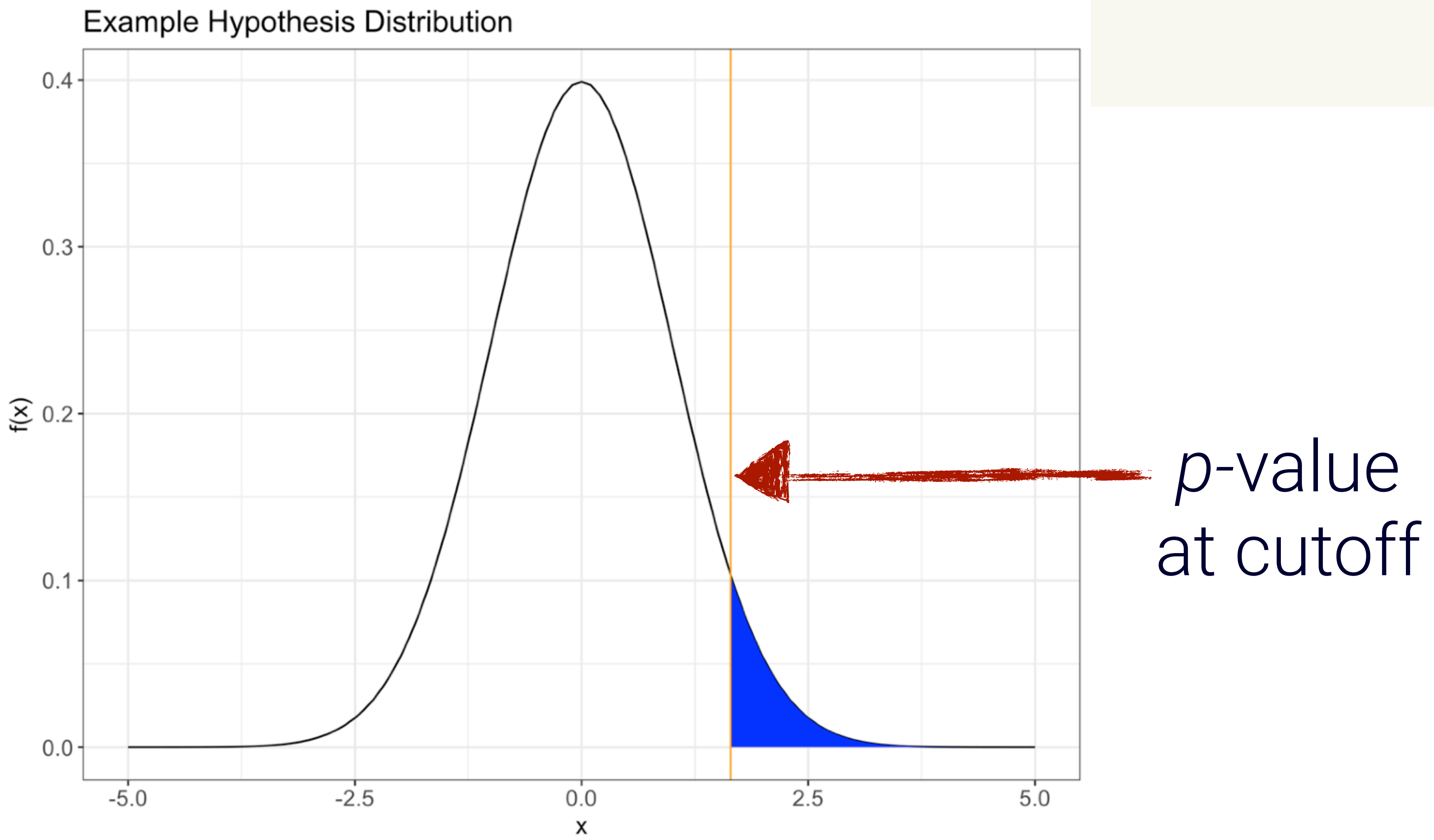
Z score

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \leq z_{1-\alpha}$$

Critical Value



Definition:
P-value is the probability associated with obtaining the observed value or more extreme values assuming the null hypothesis is true.



Toothpaste Kisses

... a story of p-values ...

Dry

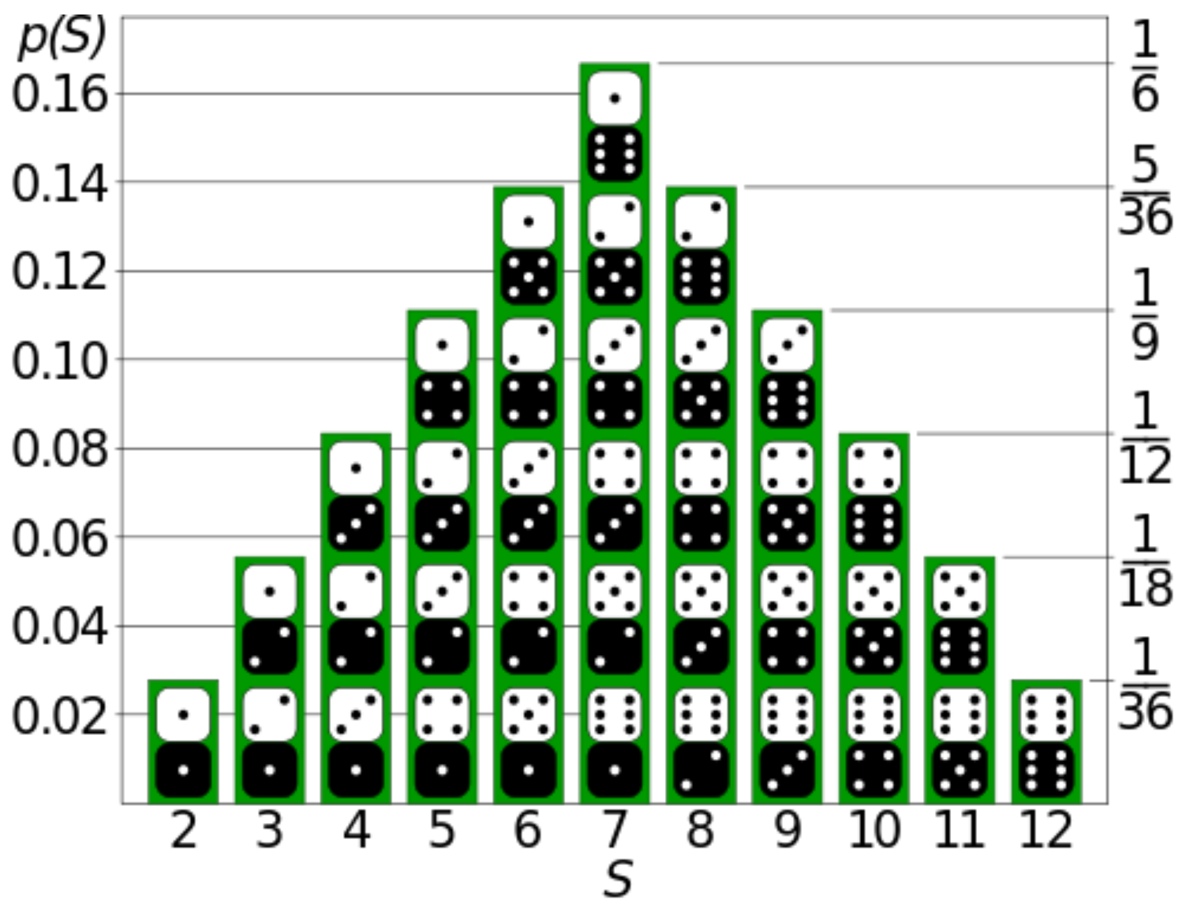
Wet

Wet + Toothpaste

Extra

How is
random phenomena
described?

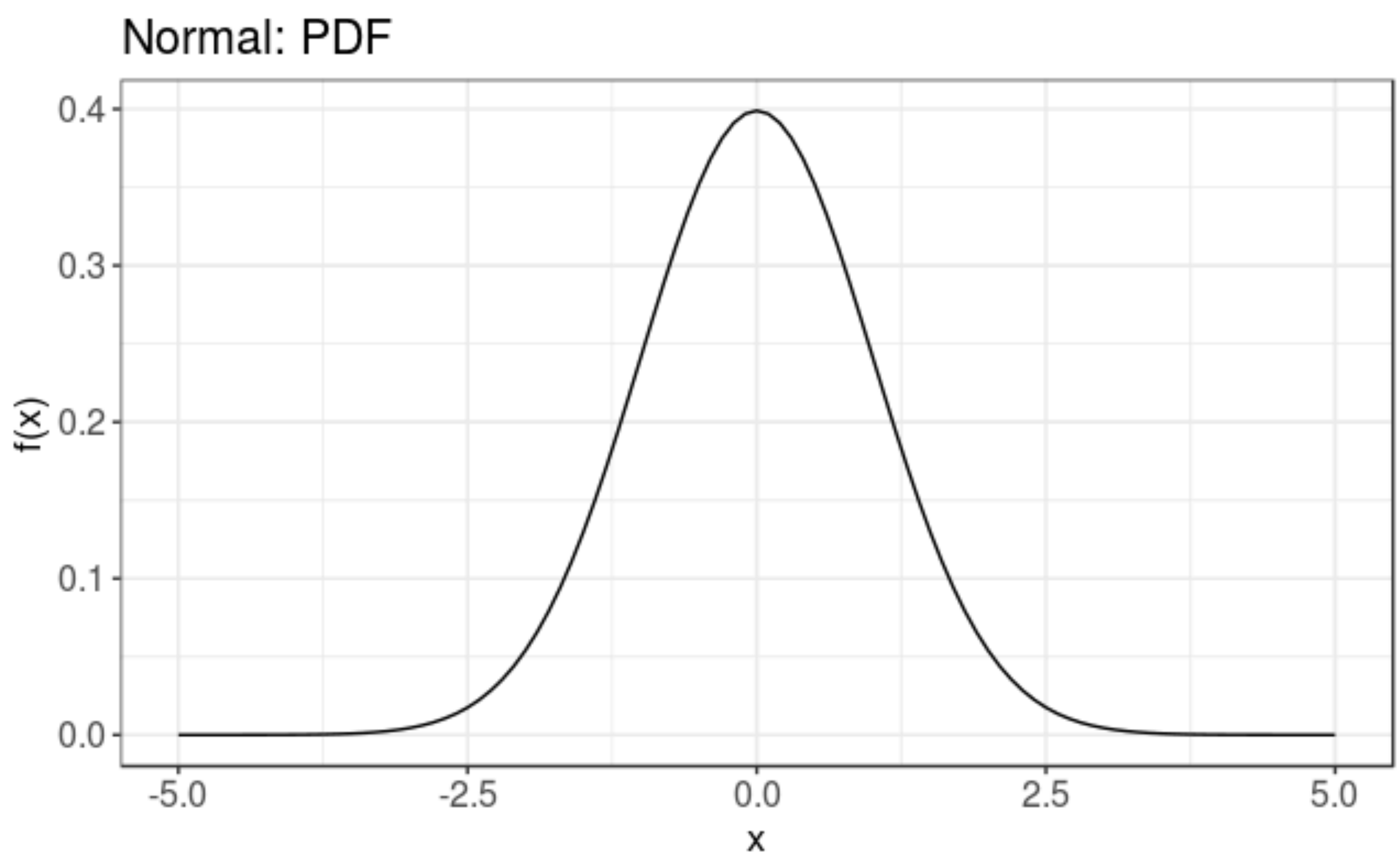
Definition:
Probability Distribution provides the outcome of a statistical event with the probability it occurs.



Source

Normal Distribution

… an example of a probability distribution …



$$f\left(x \mid \mu,\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right)$$

μ is the mean

σ is the standard deviation

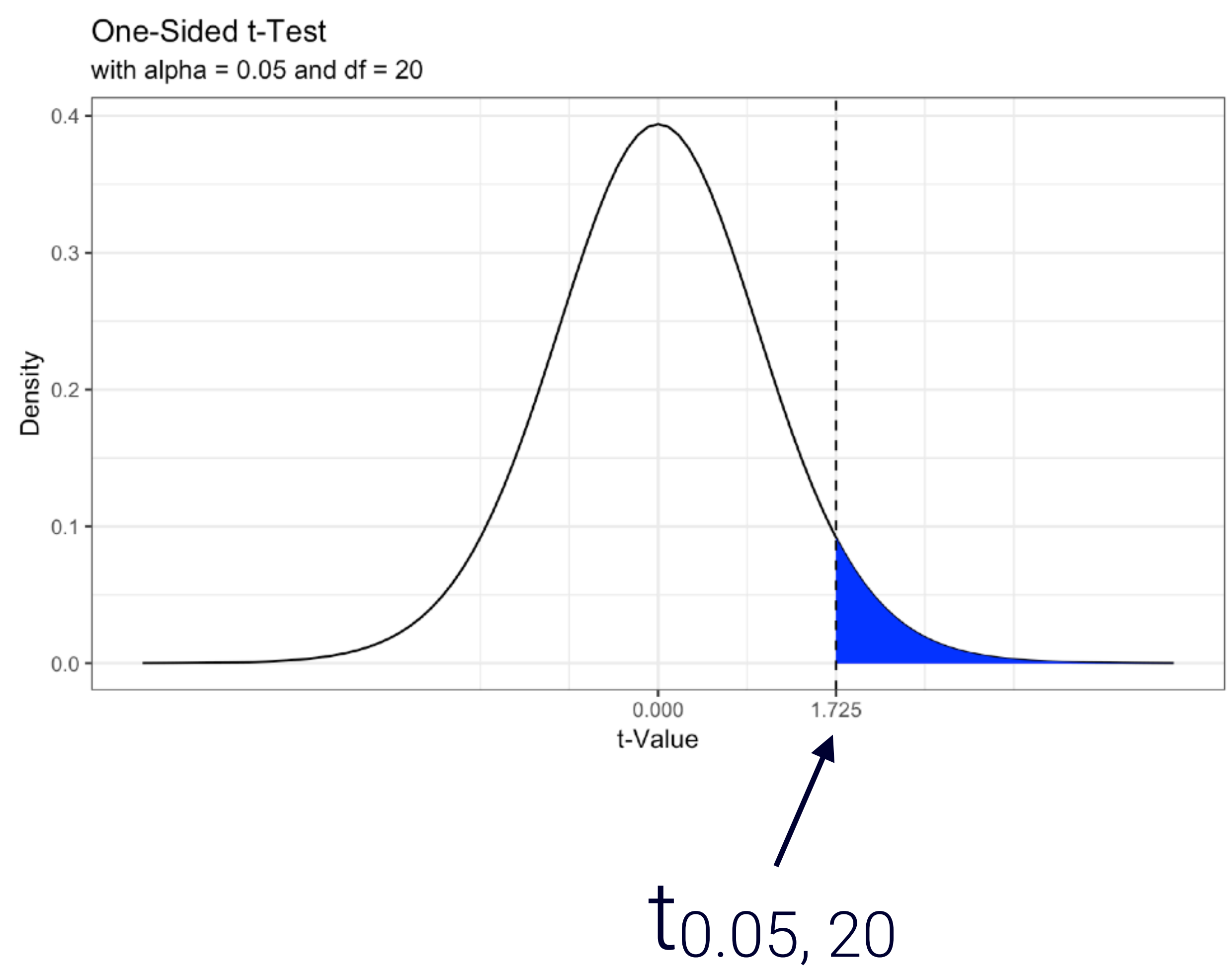
σ^2 is the variance

One-Tailed Hypothesis
t-Test

Research Question:
Is the height of males significantly *greater than* that of females?

H_0: The height of males is *<u>not</u>* significantly *greater than* females.

H_A: There is evidence to suggest the height of males *is* significantly greater than females.

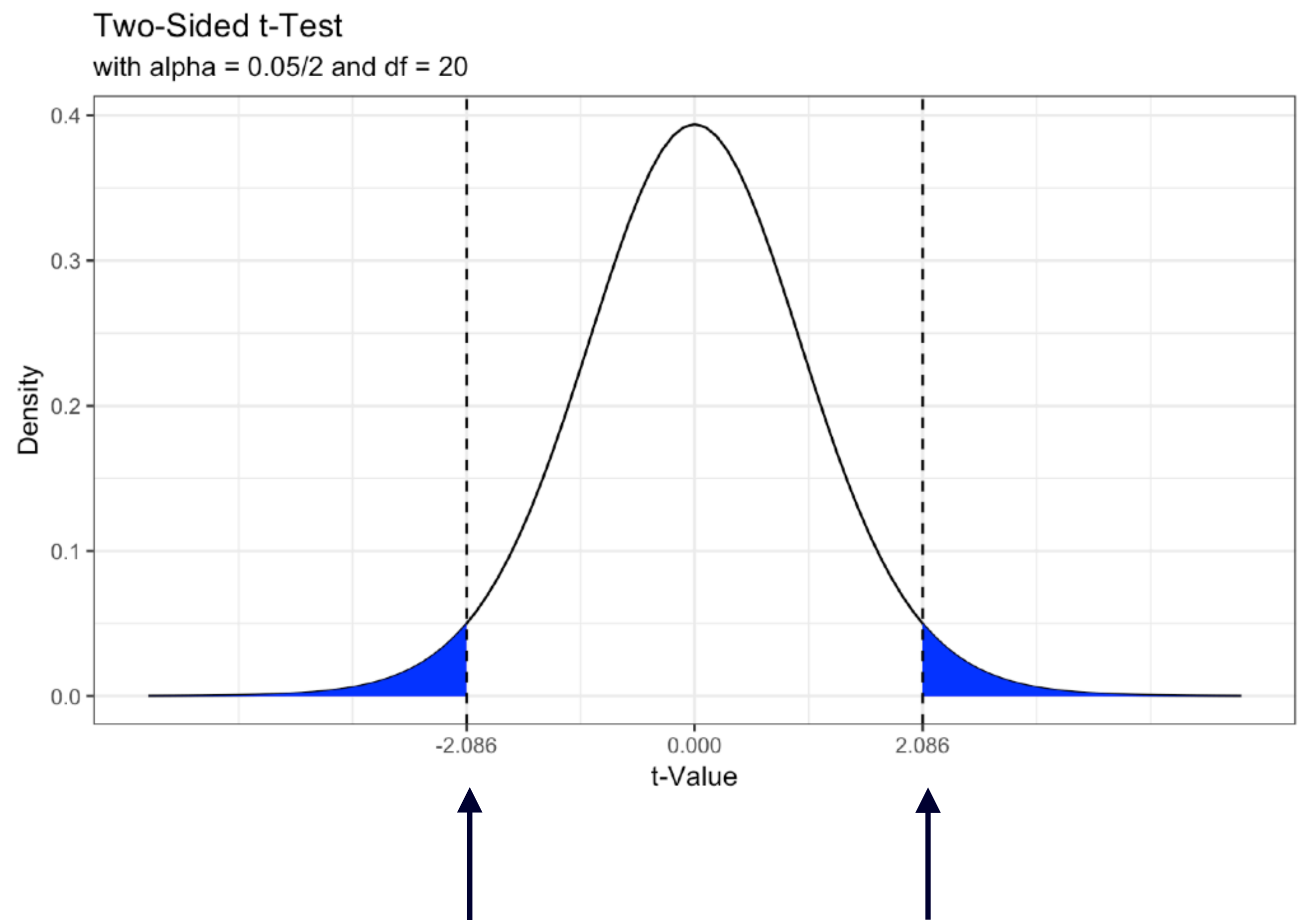


Two-Tailed Hypothesis
t-Test

Research Question:
Is there a (statistically) significant difference between the height of males and females?

H_0: There is no (statistically) significant difference between the height of males and females.

H_A: There is evidence to suggest a (statistically) significant difference between the height of males and females.



$t_{0.025, 20}$ $t_{0.975, 20}$

Summary

- **Testing Frameworks**
  - Discussed steps of a Hypothesis Test