Tutorial 09 - Hypothesis Testing

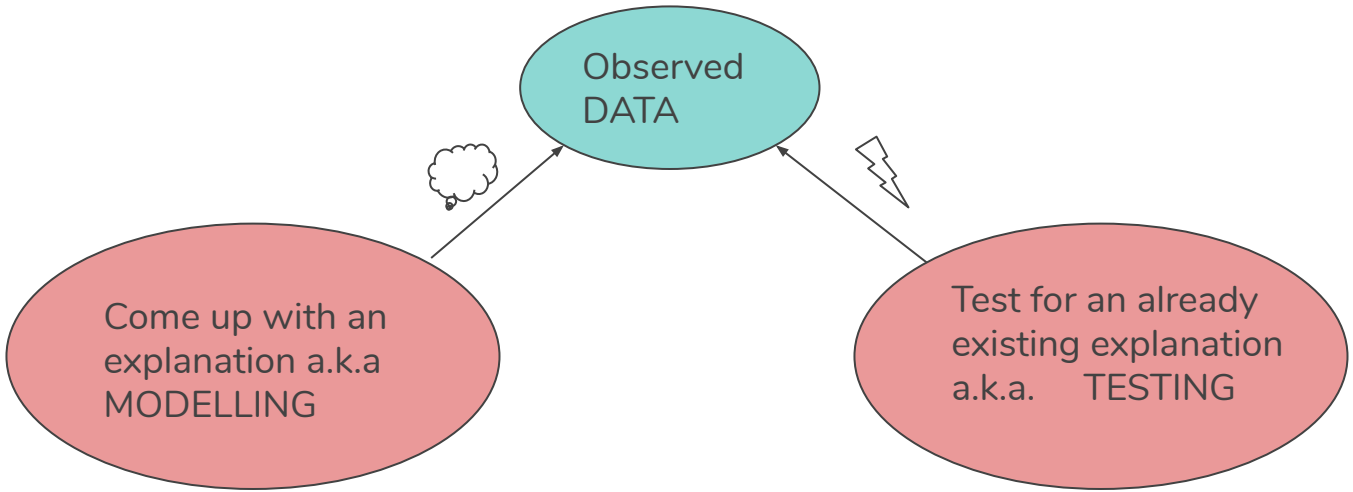
Taher & Tallulah

Table of contents

1. Revisiting *p*-value testing

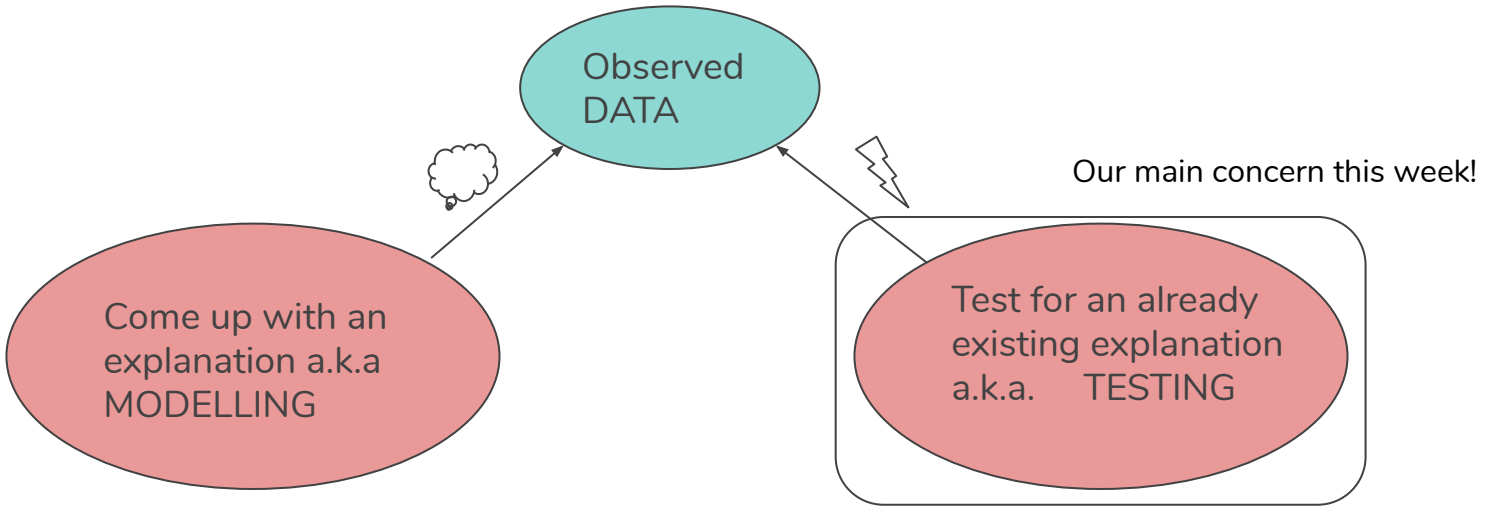
What We (don’t?) Know

Two amongst the many significant tasks for a Data Analyst:

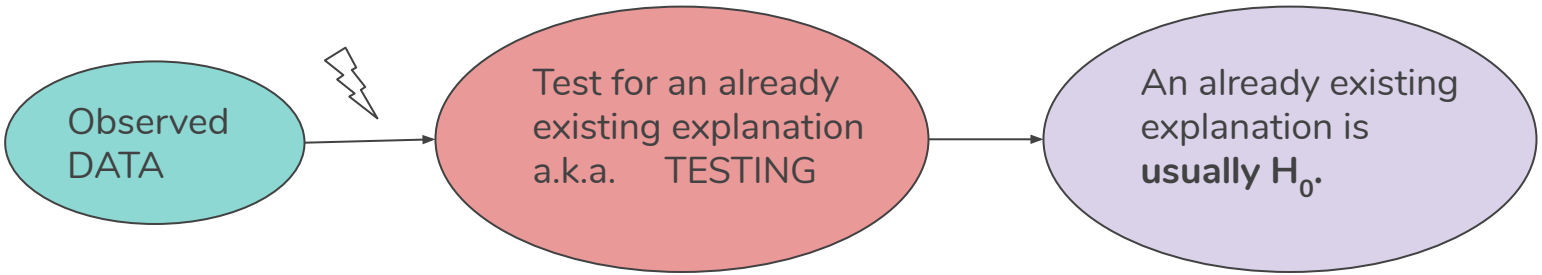


What We (don’t?) Know

Two amongst the many significant tasks for a Data Analyst:



High Level Idea of Hypothesis Testing



A Null Hypothesis H_0 is usually the one that says there is:

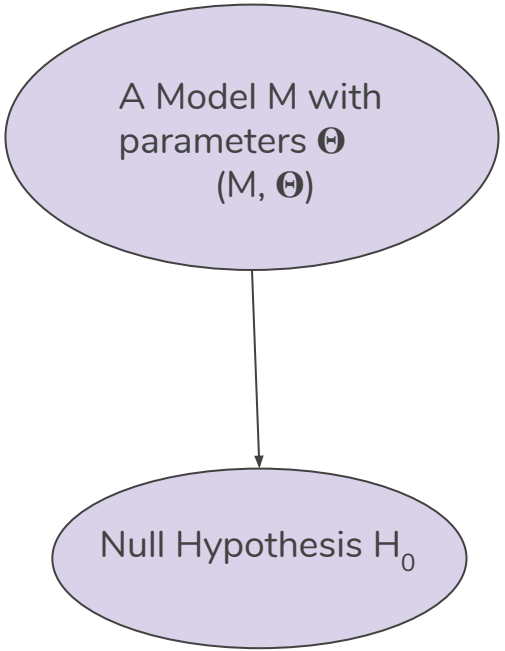
- No Effect.
- Nothing Interesting/No Surprize.
- All randomness/uncorrelated data.

High Level Idea of Hypothesis Testing

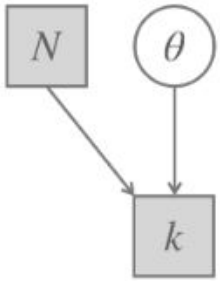
BUT:

There CANNOT be a Hypothesis regarding your observed data without an underlying model (assumption).

- However, this is not always the case! Sometimes, a model (M, Θ) is just inconceivable.

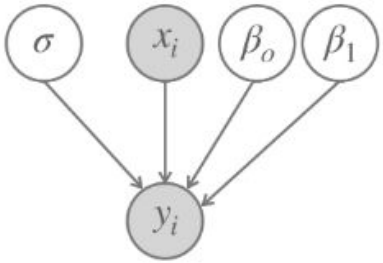


High Level Idea of Hypothesis Testing



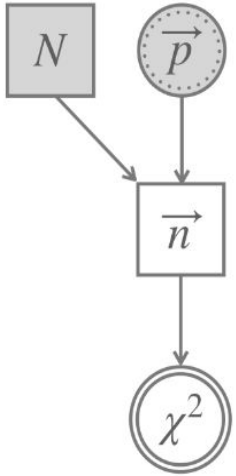
$$\theta \sim \text{Beta}(\ldots)$$
$$k \sim \text{Binomial}(\theta, N)$$

Figure 8.6: The Binomial Model



$$\sigma \sim \text{Trunc-Norm}(\ldots)$$
$$\beta_0 \sim \text{Student-t}(\ldots)$$
$$\beta_1 \sim \text{Student-t}(\ldots)$$
$$y_i \sim \text{Normal}(\beta_0 + \beta_1 x_i, \sigma)$$

Figure 8.11: The Simple Linear Regression Model.



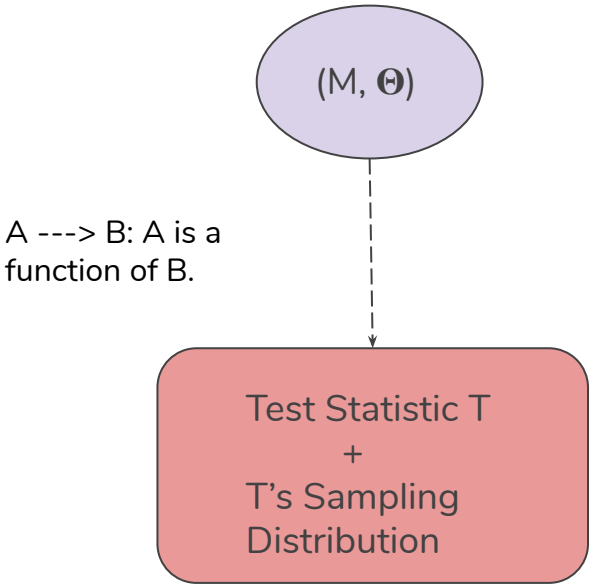
$$\vec{n} \sim \text{Multinomial}(\vec{p}, N)$$
$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

Sampling distribution:
$\chi^2 \sim \chi^2\text{-distribution}(k-1)$

Figure 10.8: Graphical representation of Pearson's χ^2 test for goodness of fit (testing a vector of predicted proportion).

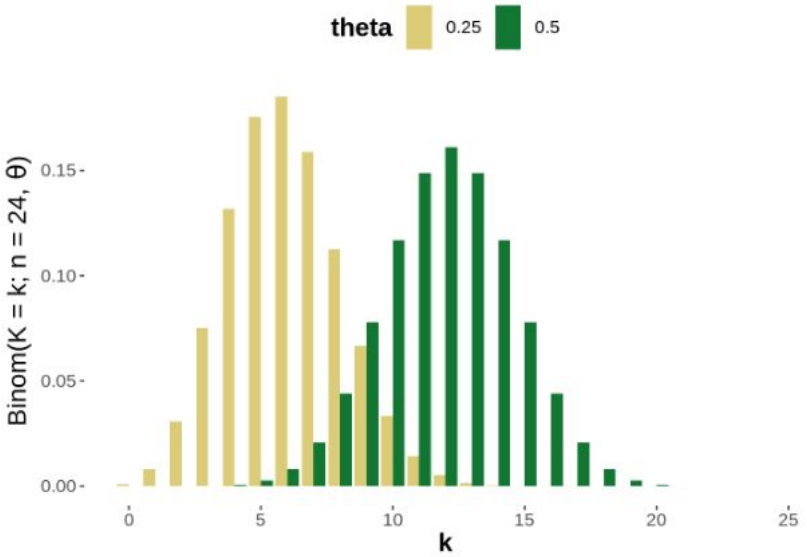
High Level Idea of Hypothesis Testing

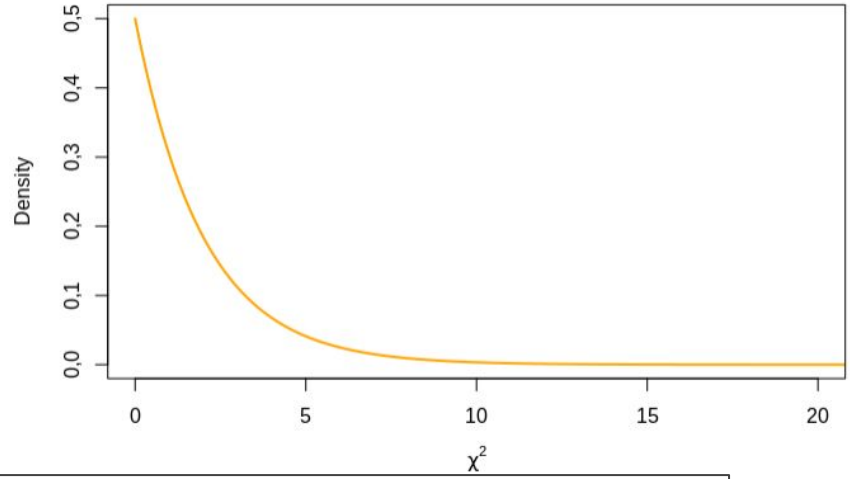
The model (M, Θ) gives us a Test Statistic and its Sampling Distribution!

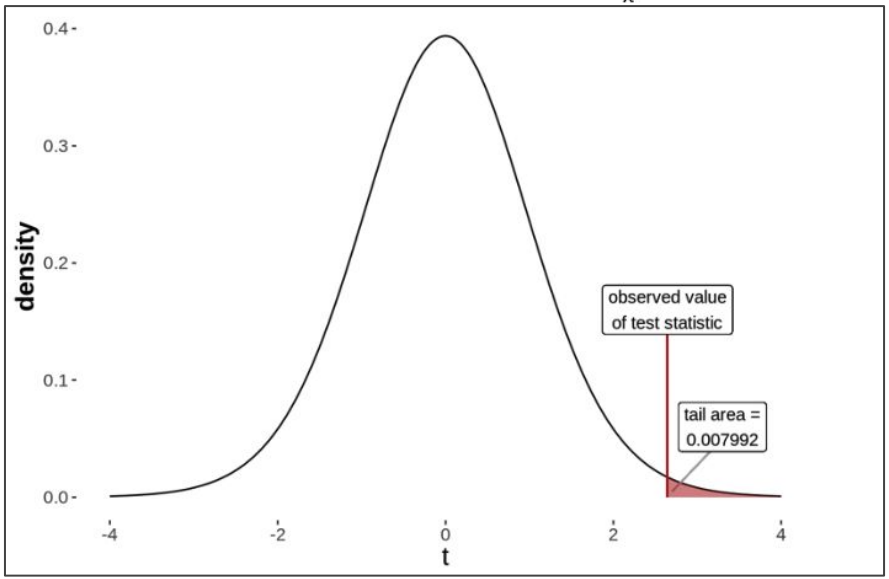


- When **we don't have an exact model** to work with ⇒ **we do not have the sampling distribution**.
- Then, we **take a large number of random samples from population to approximate the sampling distribution!**

See Exercise 4 in homework 09!

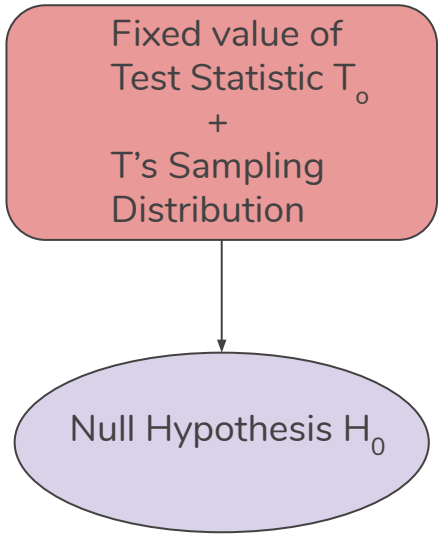

theta
0.25
0.5
Binom(K = k; n = 24, θ)
0.15
0.10
0.05
0.00
0
5
10
15
20
25
k


Density
0,5
0,4
0,3
0,2
0,1
0,0
0
5
10
15
20
χ²


density
0.4
0.3
0.2
0.1
0.0
observed value
of test statistic
tail area =
0.007992
-4
-2
0
2
4
t

High Level Idea of Hypothesis Testing

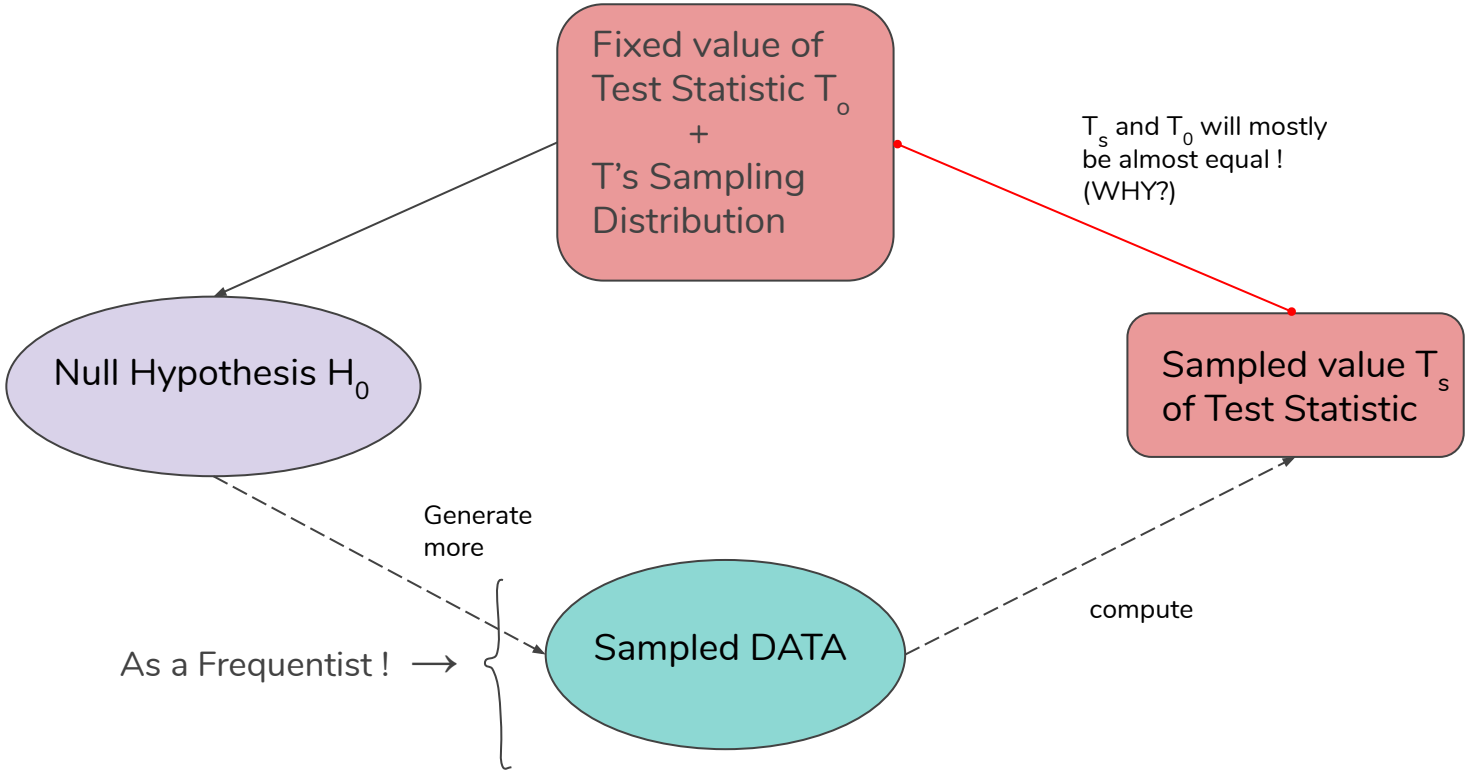
Null Hypothesis is nothing but an **assumption** that we are using a certain model (M, Θ) with a certain values of parameters Θ that gives a fixed value of test statistic T_0.



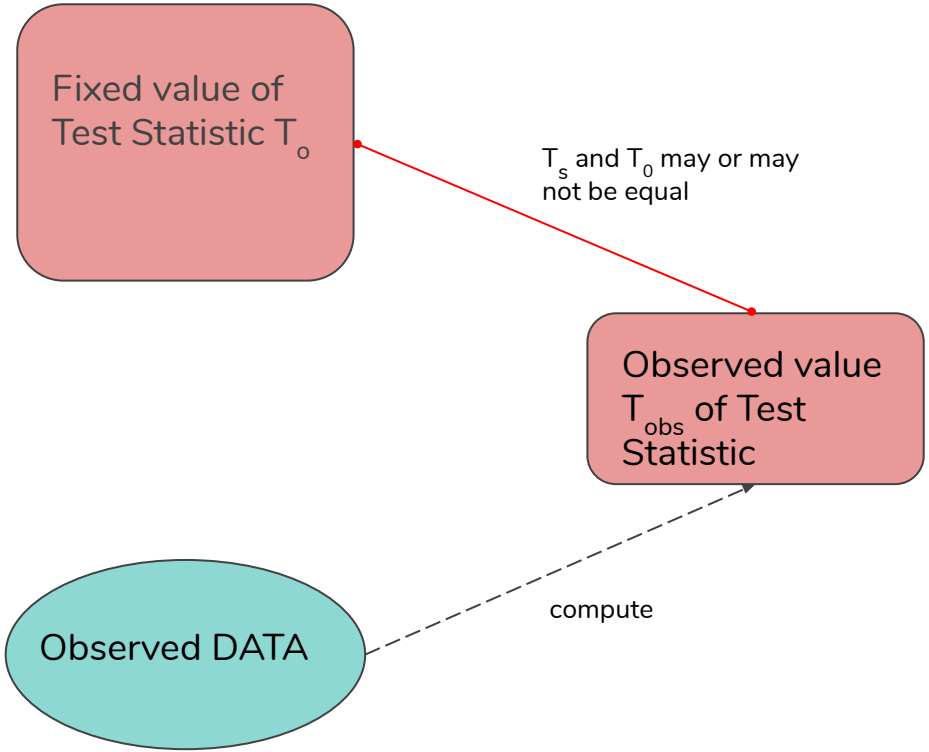
Any other different value of the test statistic can be taken as the $H_{alternate}$.

$$H_{alternate} = \neg H_0$$

High Level Idea of Hypothesis Testing



High Level Idea of Hypothesis Testing



High Level Idea of Hypothesis Testing

So What is a Test for the Model (M, θ)?

- The value of T_s depends on H_0. It will not be a fixed value but will vary since the samples are drawn from a probability distribution – the Sampling Distribution! But for large samples from the sampling distribution, T_s would be very close to T_0.
- If our Observed Data follows the H_0, then its T_{obs} should be close to T_0.
- If it is not, then it implies Observed Data is unlikely to be generated from H_0's sampling distribution.
- **The greater the difference between T_{obs} and T_o, the less likely Observed Data is to be generated from H_0's sampling distribution.**

Hypothesis Testing using *P*-value

P-value

- p-value captures the notion of **how likely Observed Data is to be generated from H_0's sampling distribution. ⇒ P(Observed Data | H_0)**
- **High** p-values ⇒ Observed Data is **more likely** to be generated from H_0.
    - ⇒ No Effect/Nothing Interesting/Nothing Surprising.
- p-values are the summation of all the probabilities of observing data that are just as much or more unlikely than the Observed Data.
- We generally choose our confidence in the test using a significance level α.
    - If p-value is lower than α, we choose to reject H_0 as the likely explanation of our data.

Hypothesis Testing using *P*-value

P-value

- p-values higher than α should be interpreted carefully.
    - If p-value is higher than α, it doesn't necessarily mean that there is no effect (i.e., H_0 is true)!
    - It only signifies that we need more data (perhaps) to reject the no-effect hypothesis. We may or may not observe such a data in real-world.
    - If p-value is lower than α, then it signifies that it is likely that there is some effect (surprise) in the observed data.

Hypothesis Testing using *P*-value

P-value

- The testing is **still not objective**, since it is upon us to choose the value of α .
- Lower the value of α, more is the control that we exercise on rejecting H_0.
    - As the value of α decreases, our demand for a very strong evidence to reject H_0 increases.

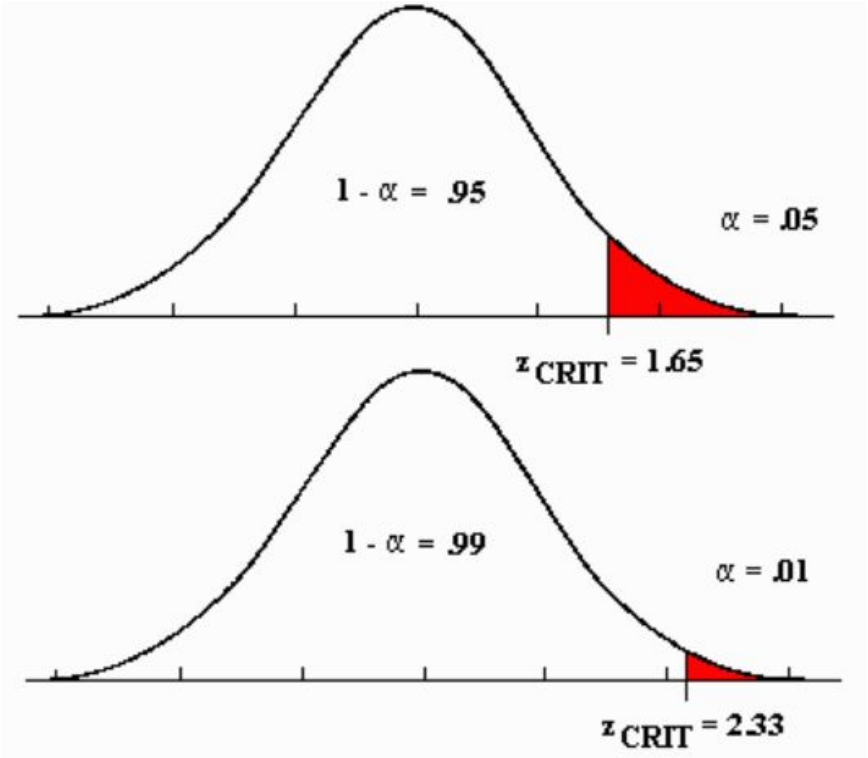


Image Source:
https://store.fmi.uni-sofia.bg/fmi/statist/education/IntroBook/sbgraph/errors01.gif

2. χ^2- test for Independence

- We have already learnt about the Pearson's chi-squared test for **goodness of fit** last week.

- Pearson's chi-squared test is also useful to **test for statistical independence of two categorical variables.**

Statistical Independence

RECAP:

Events *A* and *B* are said to be **statistically independent** iff

- $P(A \cap B) = P(A).P(B|A) = P(A).P(B)$ OR,
  - $P(A|B) = P(A)$ OR, $P(B|A) = P(B)$ ← Both of these statements are equivalent.

Intuitively,

- *Learning about the value of one, does not affect our belief in the value of the other variable.*
  - In other words, learning about the value of input variable does not provide any relevant information about the output!

Motivation

- Common to machine learning is the problem of **Feature Selection** which is about:
    - determining **what input features are relevant to predict an outcome.**
    - OR, whether a given input feature contains prediction-relevant information about the outcome.
- In a **classification task**,
    - Output → *Categorical* type
- If an input variable is also categorical, we can use the chi-squared test to determine whether the output variable is dependent on the input variable.

An Example

Two variables:

- **Sex:** Male, Female
- **Interests:** Sciences, Humanities, Commerce

	Sciences	Humanities	Commerce
Male	260	120	130
Female	200	240	50

Total Samples: 1000
Total Males: 510, Females: 490

Research Question:

Are Sex and Interests dependent?

OR,

Does the sex of a person tell us something about their interest in Science or Humanities or Commerce?

Hypotheses

H_0 : A is independent of *B*.

$H_{alternate}$: A is not independent of *B*.

Where,

$H_{alternate}$ implies that knowing the category of variable A can help you predict the category of variable B.

Caution: While support for $H_{alternate}$ suggests a relationship between *A* and *B*, it doesn't mean that the relationship is necessarily causal!

Requirements for the Test

The following conditions must be TRUE:

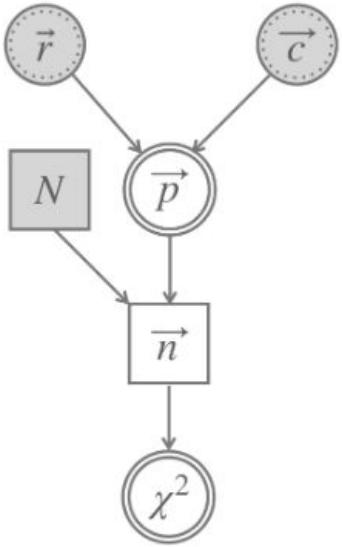
- The sampling from the population is **random.**
- The variables of the test are **categorical**.
- If the sampled categorical data are displayed in a tabular form, then **at least 80% of the cells of the table have a count of at least 5.**

The last condition is crucial for the test to be effective!

Model & Test Statistic: χ^2

- The test statistic χ^2 has a **chi-squared (sampling) distribution.**

- For our given data, chi-squared (sampling) distribution is with (2-1).(3-1)= **2 degrees of freedom!**



$\vec{p}$ = vec. of outer product $\vec{r}$ & $\vec{c}$

$\vec{n} \sim \text{Multinomial}(\vec{p}, N)$

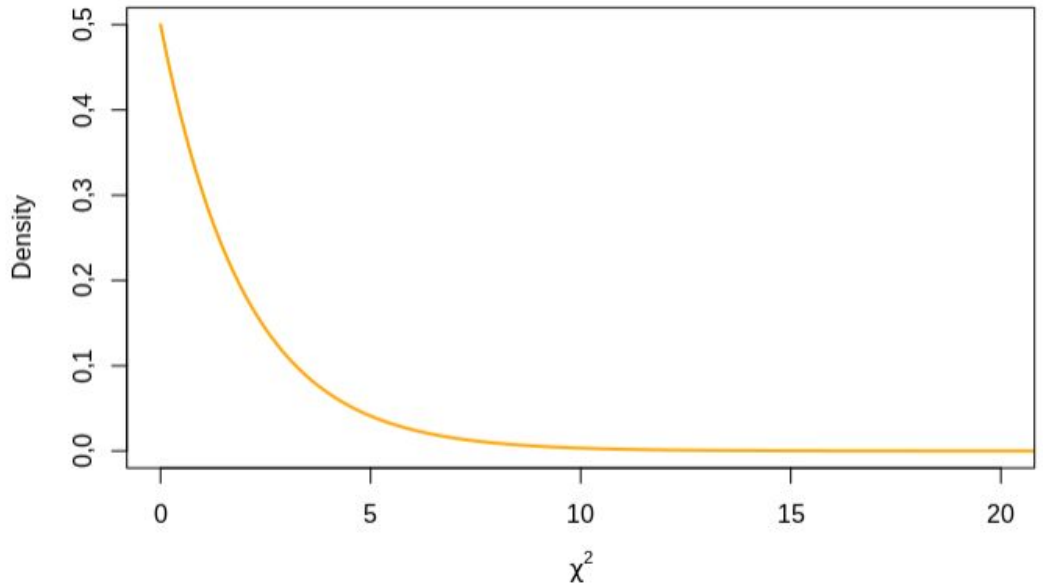
$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

Sampling distribution:

$\chi^2 \sim \chi^2\text{-distribution}\left((k_r - 1) \cdot (k_c - 1)\right)$

Test Statistic: χ^2

- A chi-squared distribution with dof=2.

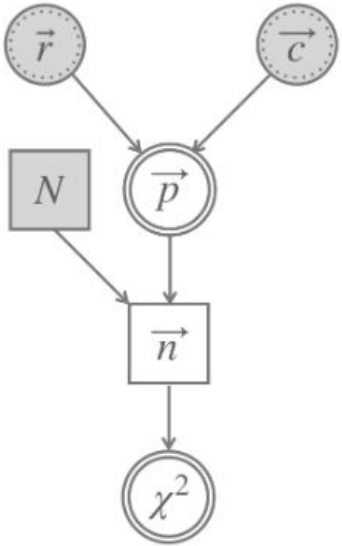


Model & Test Statistic: χ^2

- χ^2 can also be computed as:

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

where,
R and C are the number of rows and columns in the data matrix,
o_{ij} is the observed cell count in the i^{th} row and j^{th} column of the data matrix,
e_{ij} is the expected cell count in the i^{th} row and j^{th} column of the data matrix, computed as:
e_{ij} = (row i total*col j total)/total samples



$\vec{p}$ = vec. of outer product $\vec{r}$ & $\vec{c}$

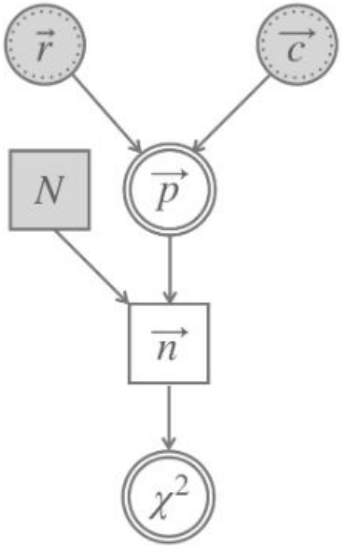
$\vec{n} \sim \text{Multinomial}(\vec{p}, N)$

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

Sampling distribution:
$\chi^2 \sim \chi^2\text{-distribution}\left((k_r - 1) \cdot (k_c - 1)\right)$

Model & Test Statistic: χ^2

- When the observed cell count is far from the expected cell count, the corresponding term in the sum is large and when the two are close, it term is small.
- So, χ^2 gives a measure of the distance between observed and expected frequencies.
- **If the two variables are really independent, then $\chi^2 = 0$. This value is the test statistic T_0 for the null hypothesis H_0.**



$\vec{p}$ = vec. of outer product $\vec{r}$ & $\vec{c}$

$\vec{n} \sim \text{Multinomial}(\vec{p}, N)$

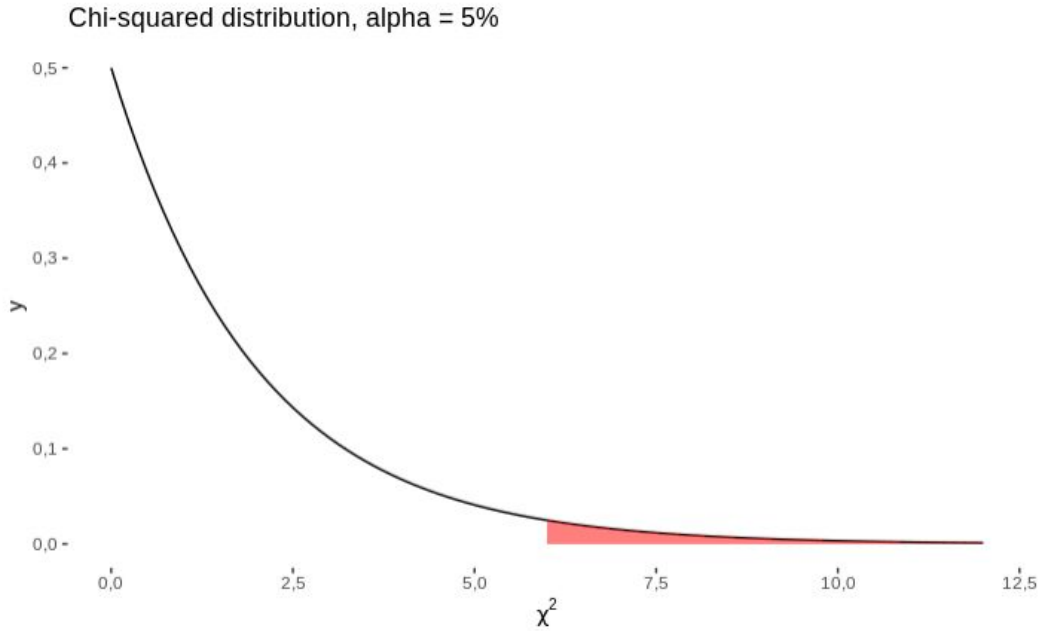
$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

Sampling distribution:

$\chi^2 \sim \chi^2\text{-distribution}\left((k_r - 1) \cdot (k_c - 1)\right)$

Test Statistic: χ^2

Higher the value of observed χ^2 ⇒ Lower would be its probability and the probability of observing more extreme values ⇒ Lower would be the p-value ⇒ more evidence <u>against</u> the H_0.



Example In R

- Creating the data matrix in R.

```r
d_vec <- c(260,120,130,200,240,50)
d_matrix <- matrix(d_vec,
                   nrow = 2,
                   byrow = T)
colnames(d_matrix) <- c('Sciences', 'Humanities', 'Commerce')
rownames(d_matrix) <- c('Males', 'Females')
d_matrix
```

- Computing all the required vectors for computing χ^2.

Probability vector of variable **Sex.**

Probability vector of variable **Interests.**

Every (*i*, *j*) entry in this matrix represents the term: P(row=*i*).P(col=*j*).N. In other words, it represents our independence assumption:

```r
# TOTAL Observations
N <- sum(d_matrix)

# vector r in the model graph
r <- d_matrix %>% rowSums() / N
# vector c in the model graph
c <- d_matrix %>% colSums() / N

# table of expected frequencies under independence assumption
d_expectation_matrix <- (r %o% c) * N

# vector p in the model graph
d_expectation_vec <- as.vector(d_expectation_matrix)
```

P(Sex=Male, Interests=Science) = P(Sex=Male).P(Interests=Science)

Example In R

- Creating the data matrix in R.

```
d_vec <- c(260,120,130,200,240,50)
d_matrix <- matrix(d_vec,
                   nrow = 2,
                   byrow = T)
colnames(d_matrix) <- c('Sciences', 'Humanities', 'Commerce')
rownames(d_matrix) <- c('Males', 'Females')
d_matrix
```

```
        Sciences Humanities Commerce
Males        260        120      130
Females      200        240       50
```

- Computing all the required vectors for computing χ^2.

```
# TOTAL Observations
N <- sum(d_matrix)

# vector r in the model graph
r <- d_matrix %>% rowSums() / N
# vector c in the model graph
c <- d_matrix %>% colSums() / N

# table of expected frequencies under independence assumption
d_expectation_matrix <- (r %o% c) * N

# vector p in the model graph
d_expectation_vec <- as.vector(d_expectation_matrix)
```

```
Males Females
 0,51    0,49
```

```
Sciences Humanities   Commerce
    0,46       0,36       0,18
```

```
        Sciences Humanities Commerce
Males      234,6      183,6     91,8
Females    225,4      176,4     88,2
```

Example In R

- Computing the test statistic χ^2.

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

= 83.01485

```
test_chi2 <- sum(
  (d_matrix - d_expectation_matrix)^2 /
   d_expectation_matrix
  )
```

- Computing the p-value.

```
p_value <- 1-pchisq(q = test_chi2_obs, df = 2)
```

```
p-value < 2,2e-16
```

Note: In computing the p-value here, we assume a chi-squared sampling distribution and the observed test statistic value is 83.01485.

Using in-built R function:

```
chisq.test(d_matrix, correct = FALSE)
```

```
        Pearson's Chi-squared test

data:  d_matrix
X-squared = 83,015, df = 2, p-value < 2,2e-16
```

Interpretation of Test

- For the given data and a chosen significance level, say at 0.05:
    - With *p*-value of almost 0, we should conclude that there is an indication of strong evidence *against* the assumption of independence (null hypothesis).

- Consequently, there is a strong evidence *in favour* of the hypothesis that: **Sex of a person can help us predict the Interests of that person.**

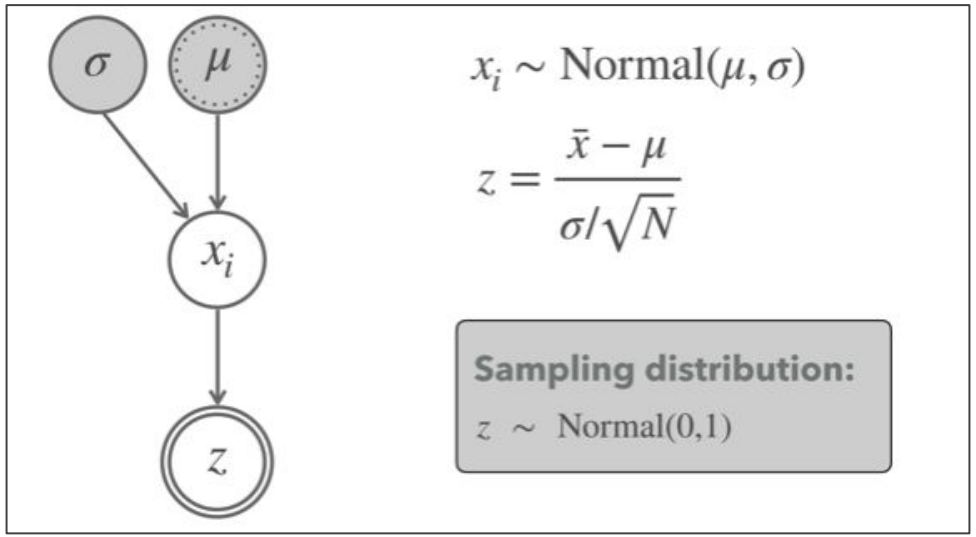
!Caution

A chi-squared test is blind to the meaning of the categories in your categorical variables. Thus, be careful when constructing your categories!

- For instance, if you are working with data about students in a university and your defined hypothesis depends on grades of the students, then you may divide the number of students into categories of those who ‘Pass’ and those who ‘Fail’ or you may also choose to divide them based on the categories of GPA (1.0-1.4, 1.5-1.9, 2.0-2.4,...) etc., the **chi-square test will treat the divisions between those categories exactly the same!**

- Thus, **it's up to you to decide whether your categories make sense,** i.e., whether the difference between GPA 1.4 and GPA 1.5 is enough to make the categories 1.0-1.4 and 1.5-1.9 meaningful.

3. z-test



If there is no difference between the sample mean and null value, the signal in the numerator, as well as the value of the entire ratio, equals zero

As difference increases, signal increases

The larger the z score, the more difference there is between samples

Assumptions: x is normally distributed and continuous, we know σ, sample mean $\bar{x}$ and the sample size N

Remember the IQ example

research hypothesis: CogSci students have, on average, a higher IQ (mean of 100 assumed for the H0)

$$H0: \mu = 100 \qquad H1: \mu > 100$$

Is it plausible to maintain that this data was generated by a normal distribution with mean 100 (if we assume that the standard deviation is known to be 15)?

```r
# number of observations
N <- length(IQ_data)
# null hypothesis to test
mu_0 <- 100
# standard deviation (known/assumed as true)
sd <- 15
z_observed <-  (mean(IQ_data) - mu_0) / (sd / sqrt(N))
z_observed %>% round(4)
## [1] 1.5802
```

We calculated the z-score, now we can calculate the p-value! (also has built-in function not shown here)

Sampling distributions assume you draw repeated random samples from a population where the null hypothesis is true

You place the z-value from your study in the z-distribution (your sampling dist) to determine how consistent your results are with the null hypothesis

```
p_value_IQ_data_ztest <- 1 - pnorm(z_observed)
p_value_IQ_data_ztest %>% round(6)
## [1] 0.057036
```

1- “pnorm” gives you the area of probabilities for observing values in this area

Z-score tells you how many sd a value is from the mean (score of z represents an element equal to the mean)

H0: μ = 100
H1: μ > 100

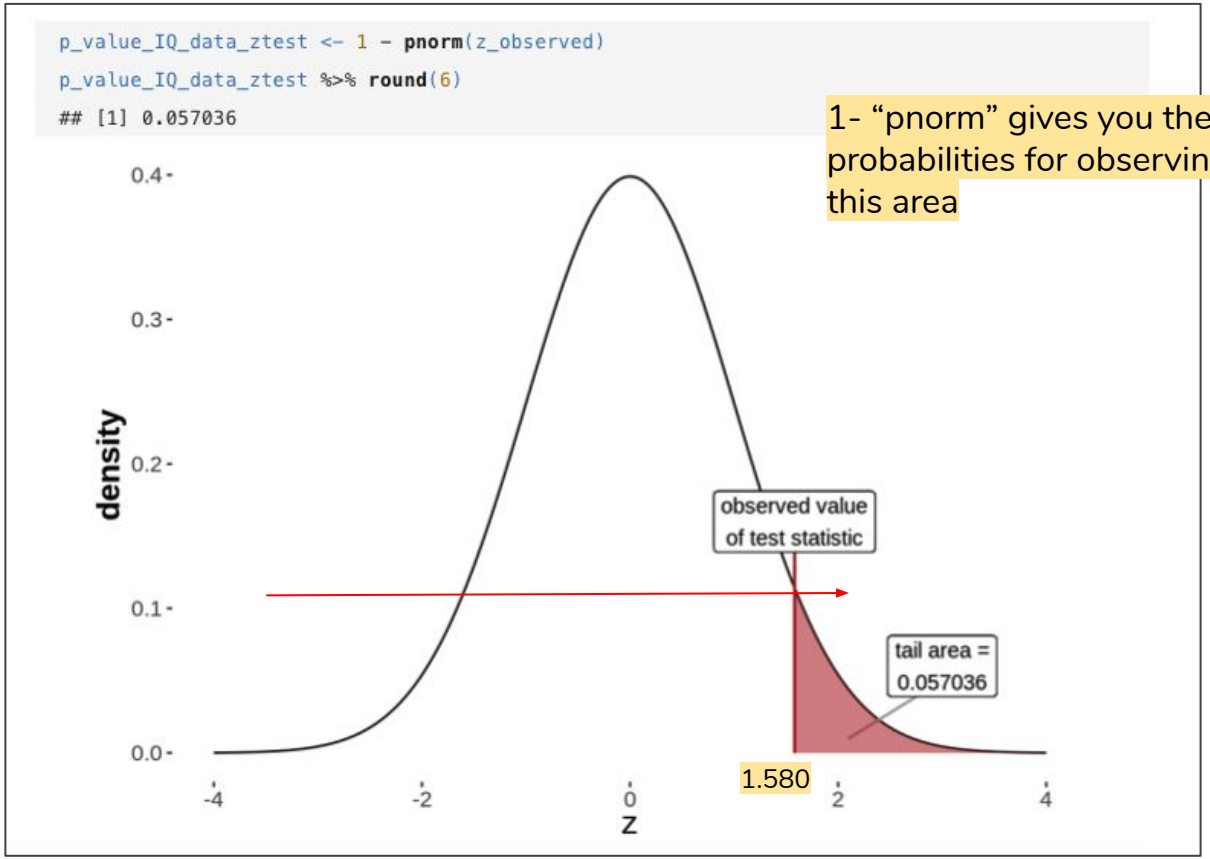
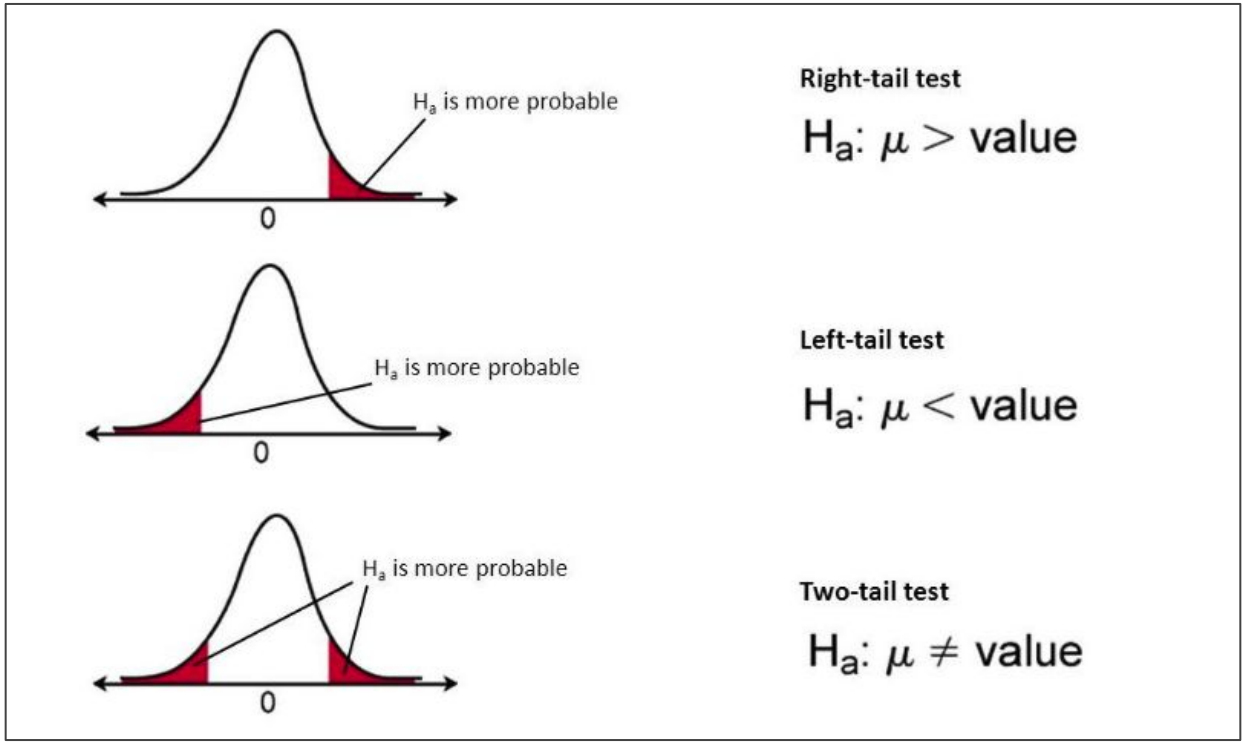


Figure 10.13: Sampling distribution for a z-test, testing the null hypothesis based that the IQ-data was generated by $\mu = 100$ (with assumed/known σ).

Cheat-Sheet


H_a is more probable
0
H_a is more probable
0
H_a is more probable
0
Right-tail test
H_a: $\mu >$ value
Left-tail test
H_a: $\mu <$ value
Two-tail test
H_a: $\mu \neq$ value

4. One-sample *t*-test

One-sample *t*-test

determines whether a **sample mean is statistically different from a hypothesized population mean**

does not assume that the standard deviation is known → uses the observed data to obtain an estimate for this parameter

we calculate a t-value

calculated t-value **is the test statistic** which we will interpret in the **context of the sampling distribution (t-distribution)**

for every t-value there is a p-value to go along with

distribution of the test statistic t is **Student's t-distribution**

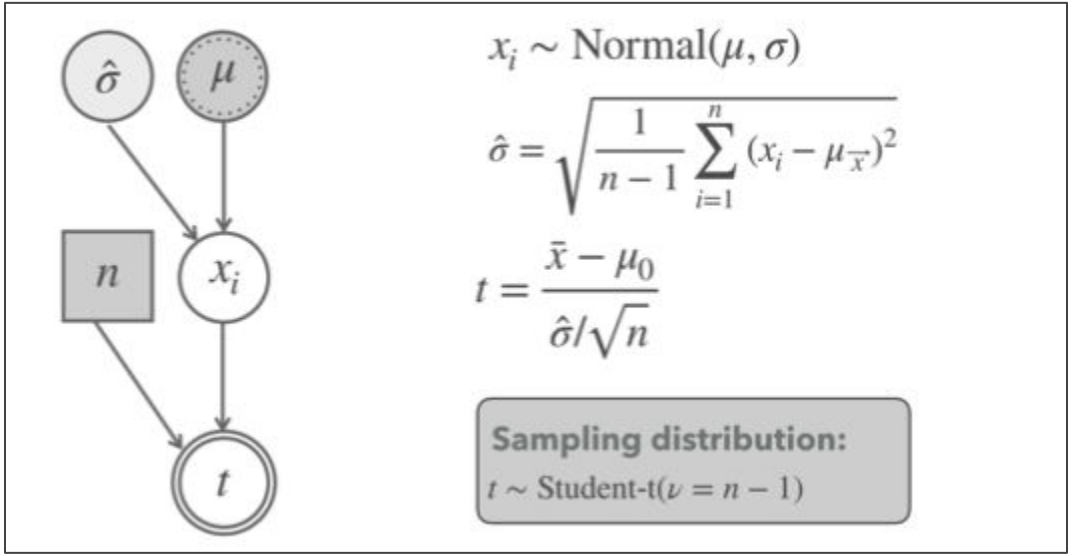


Figure 10.14: Graphical representation of the model underlying a frequentist one-sample t-test. Notice that the lightly shaded node for the standard deviation represents that the value for this parameter is estimated from the data.

If there is no difference between the sample mean and null value, the signal in the numerator, as well as the value of the entire ratio, equals zero

As difference increases, signal increases

The larger the t score, the more difference there is between samples

Back to the IQ-data Example

H0: μ = 100

H1: μ >100

one-sided p-value because our “research” hypothesis is that CogSci students have, on average, a higher IQ

```r
N <- length(IQ_data)
# fix the null hypothesis
mean_0 <- 100
# unlike in a z-test we use the sample to estimate SD
sigma_hat <- sd(IQ_data)
t_observed <- (mean(IQ_data) - mean_0) / sigma_hat * sqrt(N)
t_observed %>% round(4)
## [1] 2.6446
```

Sampling distributions assume you draw repeated random samples from a population where the null hypothesis is true

You place the t-value from your study in the t-distribution (your sampling dist) to determine how consistent your results are with the null hypothesis.

t-value of 0 indicates that the sample results exactly equal H0

As the difference between sample data & H0 increases, the absolute value of the t-value increases

We need a larger context to interpret them → t-distributions

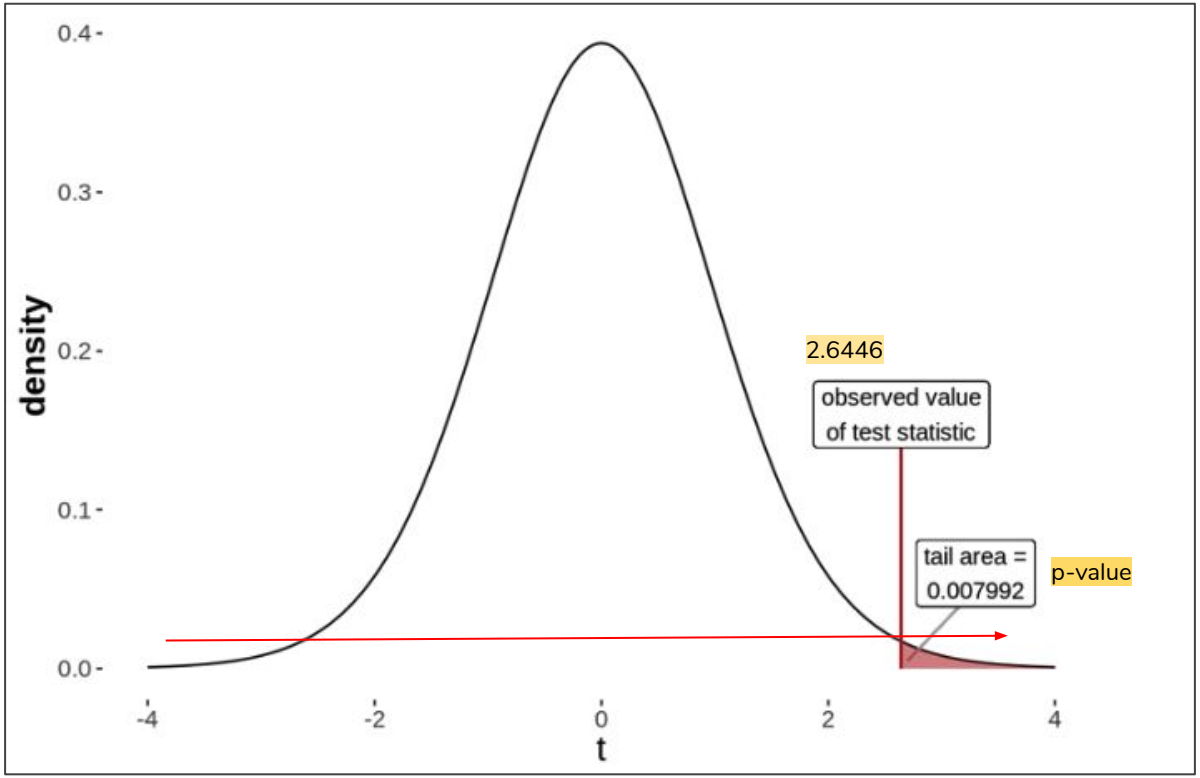


Figure 10.15: Sampling distribution for a t-test, testing the null hypothesis based that the IQ-data was generated by $\mu = 100$ (with unknown σ).

```r
p_value_t_test_IQ <- 1 - pt(t_observed, df = N-1)
p_value_t_test_IQ %>% round(6)
## [1] 0.007992
```

p-value: accumulated probabilities of observing more extreme values (less than) than what we have sampled (t-value) confined to the area of interest as indicated by our hypothesis pair (in this case, the right tail end of the curve gets considered as the probability of observing values there would be increasingly unlikely towards the tail end and therefore indicate that H0 cannot be true for the population and cannot have generated the sampled data)

```r
t.test(x = IQ_data, mu = 100, alternative = "greater")
##
##  One Sample t-test
##
## data:  IQ_data
## t = 2.6446, df = 19, p-value = 0.007992
## alternative hypothesis: true mean is greater than 100
## 95 percent confidence interval:
##  101.8347      Inf
## sample estimates:
## mean of x
##     105.3
```

One-sample *t*-test two-tailed

Example :

We know the trait “neuroticism” is normally distributed in the population. We assume the population mean of this distribution as a scoring of 2.5 on average on the big-five scale. We are interested in whether in reality, the true mean is either above or below the assumed population average based on our sample mean.

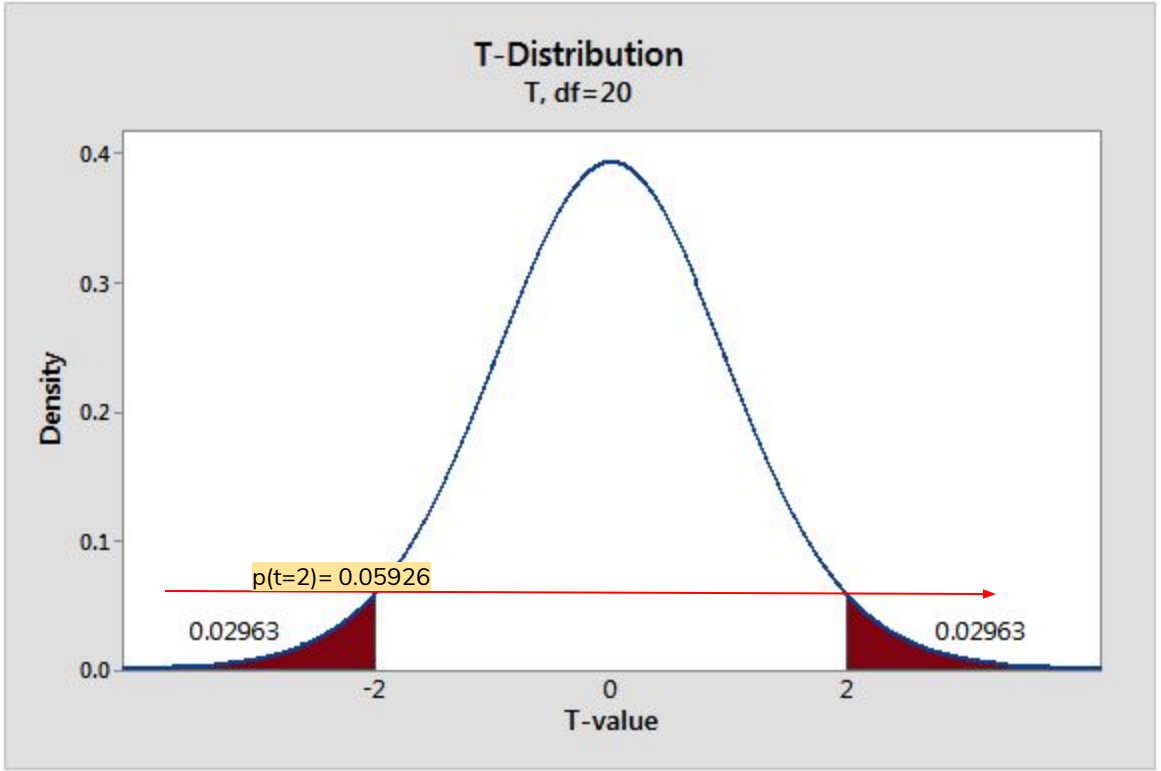
$$H0: \mu = 2 \quad H1: \mu \neq 2$$

Compute the t-value → its 2!

In the next picture: probability associated with t-values less than -2 and greater than +2 using the area under the curve because we are interested in both

each shaded region has a probability of 0.02963, accumulating to 0.05926

shows that t-values fall within these areas almost 6% of the time when H0 is true!



The accumulated probability is the p-value

Not quite enough to reject H0 under the standard significance level of 0.05.

Source: https://statisticsbyjim.com/hypothesis-testing/t-tests-t-values-t-distributions-probabilities/

5. Two-sample *t*-test

Two-sample *t*-test

analyze the **difference between the means of two independent samples**

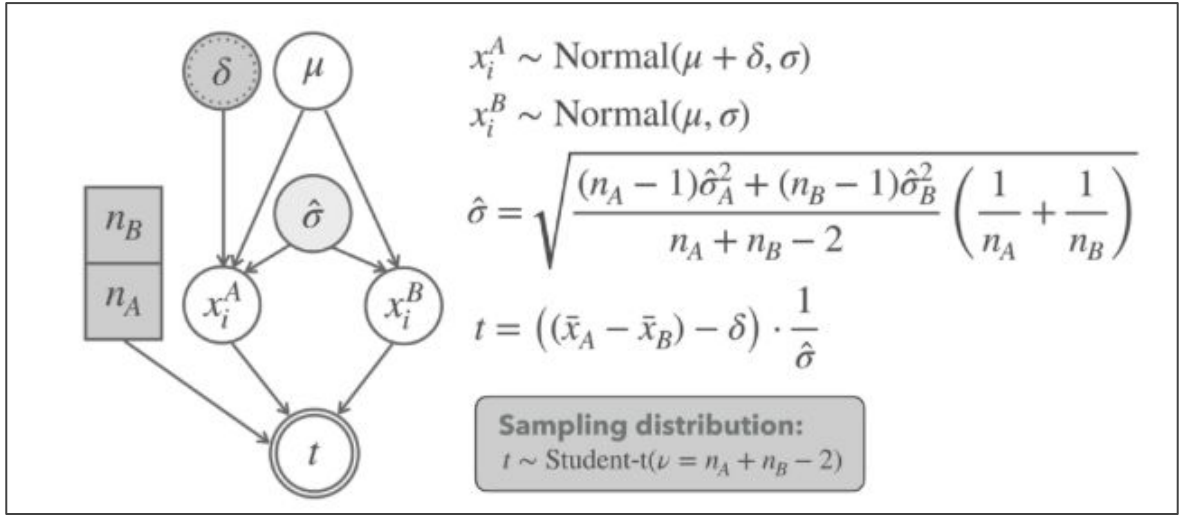
tests whether two samples are drawn from populations with different means, tests whether the underlying populations for the two samples actually differs

tells you how significant the differences between two samples is

In other words, whether a difference between two sample averages is unlikely to have occurred because of random chance in sample selection

Here, we focus on **unpaired data** (as from a between-subjects design), **assume equal variance but (possibly) unequal sample sizes**

we will proceed as usual!



The larger the t score, the more difference there is between samples

XA and XB are the price measures for conventionally grown and for organically grown avocados. Assumes XA and XB are iid samples from a normal distribution

the mean of one sample XB is assumed to be some unknown μ (it will cancel out though)

additive parameter δ indicating the difference between means of these sample (mean for XA is calculated with added δ)

Two-sample *t*-test one-tailed

Example:

investigate whether the weekly average price of **organically grown avocados is higher** than that of **conventionally grown avocados**

XA = organically grown avocados

XB = conventionally grown avocados

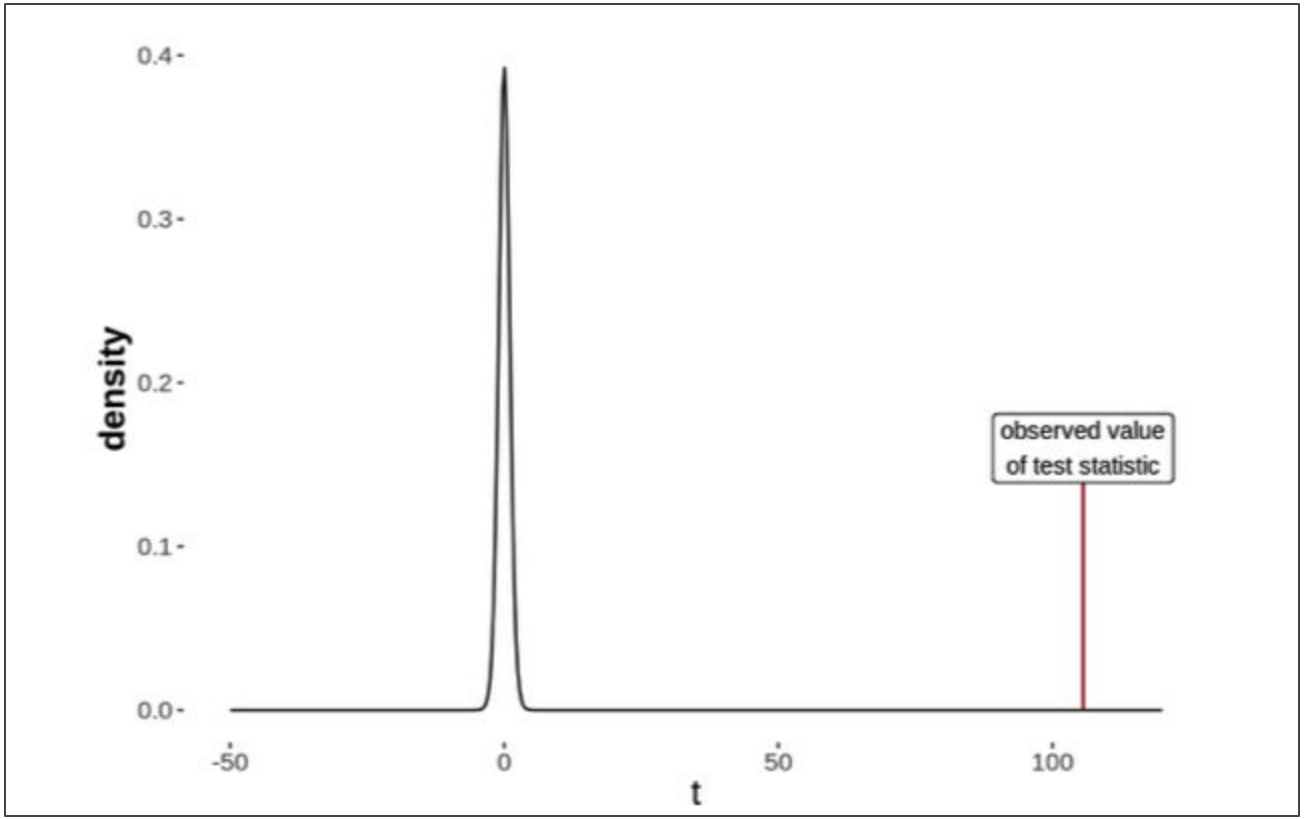
$$H0: \mu d = \mu 0 \quad H1: \mu d > 0$$

we use a **one-sided** test because we **hypothesize that organically grown avocados are more expensive,** not just that they have a different price (more expensive or cheaper)

Put computed t-value in the context of t-distribution to assess its p-value

```r
# fix the null hypothesis: no difference between groups
delta_0 <- 0
# data (group A)
x_A <- avocado_data %>%
  filter(type == "organic") %>% pull(average_price)
# data (group B)
x_B <- avocado_data %>%
  filter(type == "conventional") %>% pull(average_price)
# sample mean for organic (group A)
mu_A <- mean(x_A)
# sample mean for conventional (group B)
mu_B <- mean(x_B)
# numbers of observations
n_A <- length(x_A)
n_B <- length(x_B)
# variance estimate
sigma_AB <- sqrt(
  ( ((n_A -1) * sd(x_A)^2 + (n_B -1) * sd(x_B)^2 ) /
      (n_A + n_B -2) ) * (1/n_A + 1/n_B)
)
t_observed <- (mu_A - mu_B - delta_0) / sigma_AB
t_observed
## [1] 105.5878
```

sampling distribution for a two-sample t-test, testing the null hypothesis of no difference between samples



2.2e-16 = 2.2 to the power of -16

```
t.test(
  x = x_A,           # first vector of data measurements
  y = x_B,           # sec vector of data measurements
  paired = FALSE,    # measurements are to be treated as unpaired
  var.equal = TRUE,  # we assum equal variance in both groups
  mu = 0             # NH is delta = 0 (name 'mu' is misleading!)
)
##
##  Two Sample t-test
##
## data:  x_A and x_B
## t = 105.59, df = 18247, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4867522 0.5051658
## sample estimates:
## mean of x mean of y
##  1.653999  1.158040
```

this is a highly significant p-value indicating that just as hypothesized, organically grown are more expensive than conventionally grown avocados

Three main types of t-test:

- Two-samples t-test compares the means for two samples
- One sample t-test tests the mean of a single sample against a known or hypothesized mean
- Paired sample t-test compares means from the same sample at different times (e.g. in 10 minute intervalls)

paired means your samples are dependent, e.g. when you obtain two measures on the same participants or object under different conditions (measure body weight after one week,two weeks...)

For paired t-tests: H0 = the *pairwise difference* (all pairs from a possible set) between the two samples is equal (H0: $\mu d = 0$)

paired two-sample t-tests require the two samples to be of equal lengths (while **unpaired** two-sample t-tests do **not** require the two samples to be of equal lengths!)

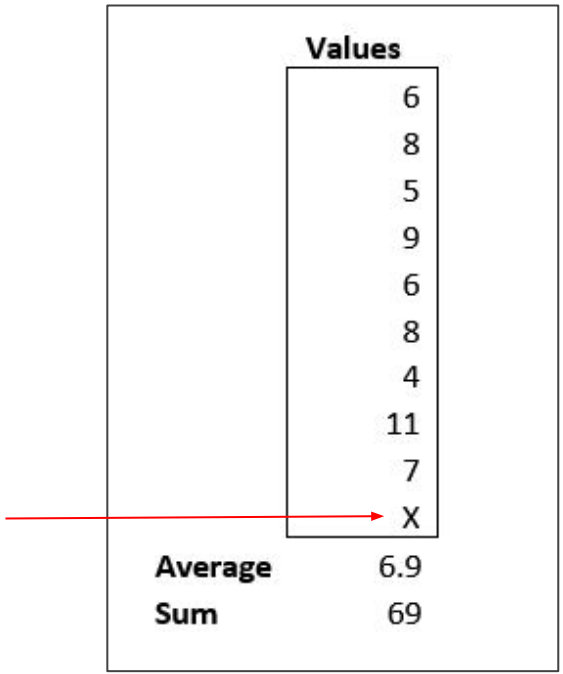
Degrees of Freedom

Student’s t-distribution works with DF

DF are the number of observations in a sample that are free to vary while estimating statistical parameters

DF are the amount of independent data that you can use to estimate a parameter

Example



We have a random sample of observations. Imagine we know the mean but we don't know the value of an observation

The mean is 6.9 based on 10 values. We know that the values must sum to 69 based on the equation for the mean

64 + X = 69 we know that X must equal 5 → last number has no freedom to vary

not an independent piece of information because it cannot be any other value

Estimating the parameter, the mean in this case, imposes a constraint on the freedom to vary

The last value and the mean are dependent on each other → after estimating the mean, we have only 9 independent pieces of information even though the sample size is 10

Degrees of Freedom

DF also define **probability distributions** for the **test statistics** of various hypothesis tests

They define e.g. t-distribution, F-distribution, and the chi-square distribution to determine statistical significance

Each these distributions is a family of distributions where the **DF define the shape**

when you have a sample and estimate the mean, you have n – 1 degrees of freedom, where n is the sample size

→ **for a 1-sample t-test, the degrees of freedom is n – 1**

→ **for a 2-sample t-test, the degrees of freedom is n1 + n2 – 2**

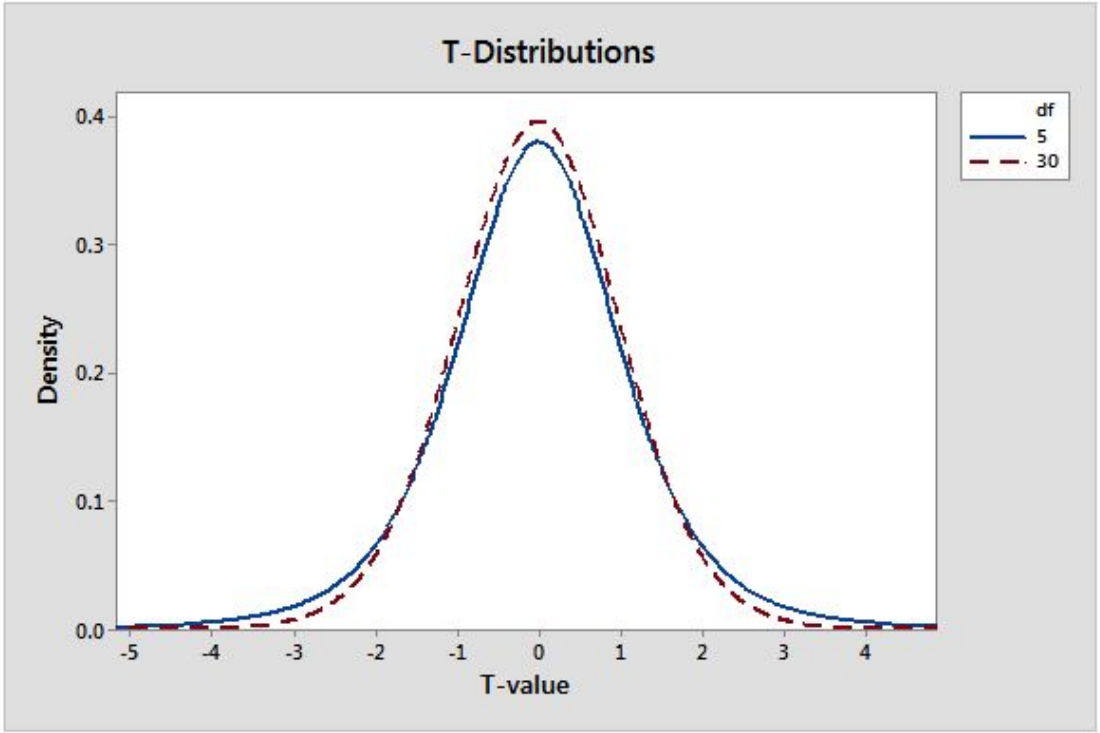
shows t-distribution for several different DF

DF define the shape

Because DF are so closely related to sample size, you can see the effect of sample size

DF decrease → t-distribution has broader tails → allows for the greater uncertainty associated with small sample sizes!

As sample size increases, the sample more closely approximates the population (normal distribution)



Source: https://statisticsbyjim.com/hypothesis-testing/t-tests-t-values-t-distributions-probabilities/

References

- Urdan T. (2010). *Statistics In Plain English*, 3rd edition.
- Kent State University n.d., *SPSS Tutorials: Chi-Squared Test for Independence*, accessed on 19th January, <https://libguides.library.kent.edu/SPSS/ChiSquare>

Questions?