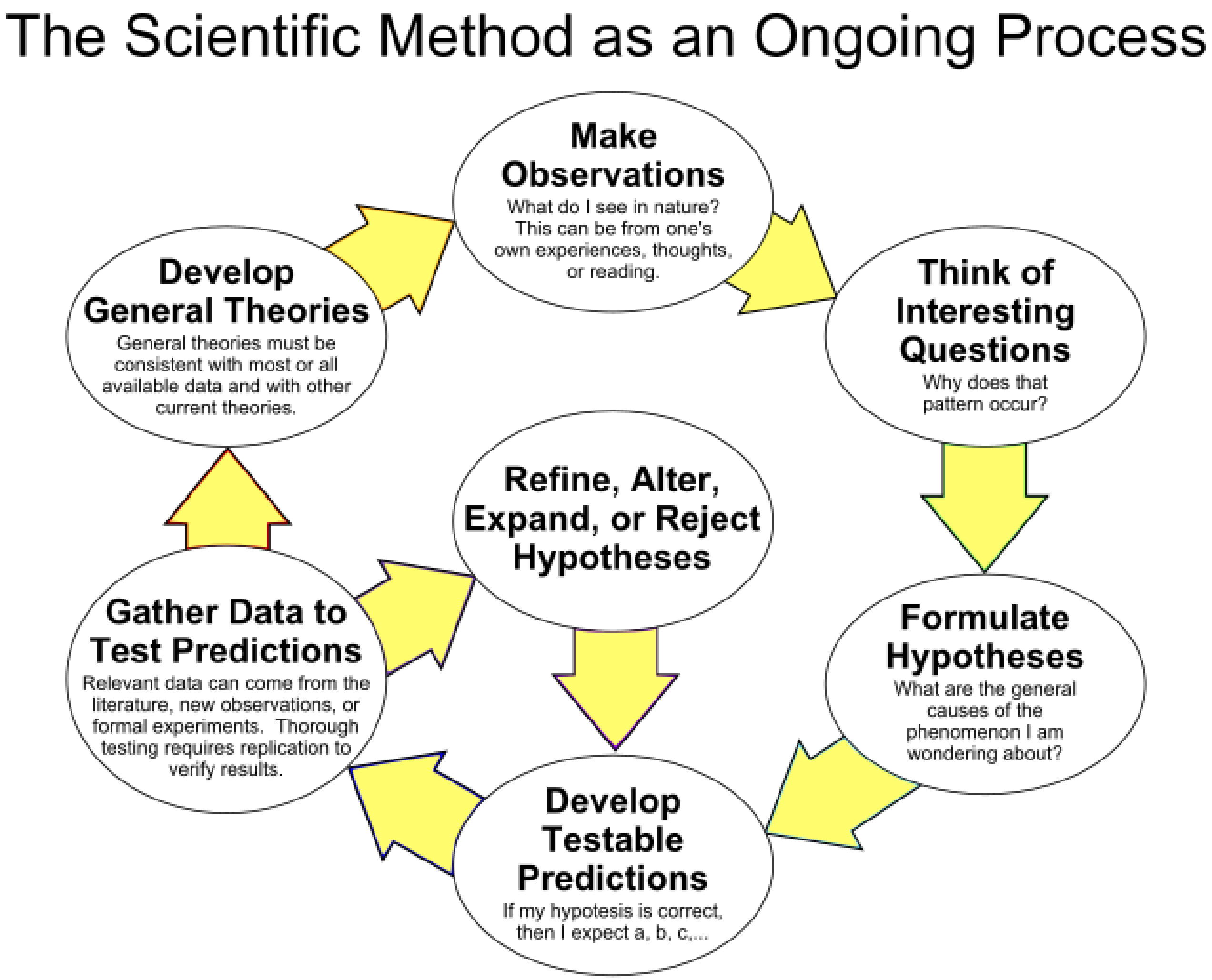
Hypothesis Testing



Given the outcome of one or several observations, we must determine whether the predicted (model) and measured values are compatible.

If the values are <u>compatible</u>, the model is considered provisionally acceptable and may be considered the main/leading model until another measurement invalidates its predictions.

If the values are <u>NOT compatible</u>, the model may be <u>rejected</u> or require further revisions.

This raises the question of how one can assess whether two values, a measurement and a prediction, or even two distinct measurements, can be considered compatible!

Hypothesis Testing

One cannot expect that the model value and the measurement match perfectly! The model may consider approximations that do not “perfectly copy” real-world phenomena, as well as experiments, and are limited in statistical accuracy and errors/uncertainties.

At best, one can hope that the difference between measured and predicted value is small, but how small? Or how significant is this difference?

One needs to quantify how large the difference can be, considering uncertainties in the measurement as well as in the formulation of the model that resulted in the prediction.

One may wish to assess whether the model has a large likelihood of being valid given the measured values and their estimated errors.

SHOULD THE MODEL BE ACCEPTED (PROVISIONALLY) OR REJECTED?

Hypothesis Testing

A statistical test is required to determine whether the measured and the predicted values are compatible. **Methods** for the formulation of such tests and their properties are within the field of **hypothesis testing**.

The formulation of a scientific test begins with a statement about the expected outcome of a measurement based on a specific model. The statement constitutes the hypothesis to be tested and is commonly referred to as the **null hypothesis**.

H_0: The **null hypothesis** ➔ a statement of no difference between the variables. This can often be considered the *status quo* (e.g., no need for *new* physics). As a result, if you cannot accept the null in whatever you are investigating, it may require some *action*.

The notion of null hypothesis can be seen as counterintuitive; the point is that **it is not possible to scientifically prove that a theory is correct for all possible circumstances**; one can demonstrate that it is incorrect or that a model is wrong or incomplete (*principle of falsifiability*).

The null hypothesis typically adopts the point of view that an existing model (or theory) is correct, but measurements can be used to falsify the model.

https://openstax.org/books/introductory-business-statistics/pages/9-1-null-and-alternative-hypotheses

Hypothesis Testing

H_a: The **alternative hypothesis** ➔ usually a claim that is contradictory to H_0 and what we conclude when we cannot accept the null. This is usually what the researcher is trying to prove. The alternative hypothesis is the contender and should win with significant evidence to overthrow the status quo.

One can define more than one alternative hypothesis.

Since the null and alternative hypotheses are contradictory, you must examine evidence to decide if you have enough to reject the null hypothesis or not. The evidence is usually in the form of sample data. After you have determined which hypothesis the sample supports, you make a **decision**. You may say "cannot accept H_0" if the sample information favours the alternative hypothesis or "do not reject H_0" or "decline to reject H_0" if the sample information is insufficient to reject the null hypothesis. These conclusions are all based upon a level of probability, a significance level, that we can set.

Statistical tests may be used in a variety of contexts and for various purposes and can be formulated in many different ways.

Hypothesis Testing

It is important to remember that hypotheses are statements **about the population** or distribution under study, not statements about the sample.

The value of the population parameter specified in the null hypothesis is usually determined in one of three ways:

(i) It may result from past experience or knowledge of the process, or even from previous tests or experiments. The objective of hypothesis testing, then, is usually to determine whether the parameter value has changed.

(ii) It may be determined from some theory or model regarding the process under study. Here the objective of hypothesis testing is to verify the theory or model.

(iii) A third situation arises when the value of the population parameter results from external considerations, such as design or engineering specifications, or from contractual obligations. In this situation, the usual objective of hypothesis testing is conformance testing.

Applied Statistics and Probability for Engineers, 5th Edition
George C. Runger, Douglas C. Montgomery

Hypothesis Testing

Various hypotheses types. H_0 and H_a are mutually exclusive and all inclusive; they cover all possibilities and when one is TRUE, the other is FALSE and vice-versa. For example, if the null hypothesis is equal to some value, the alternative has to be not equal to that value.

H_0	H_a
$=$	$\neq$
$\geq$	$<$
$\leq$	$>$

When we perform a hypothesis test, there are four possible outcomes depending on the actual truth (or falseness) of the null hypothesis H_0 and the decision to reject or not. The outcomes are summarized below

Statistical Decision	H_0 is actually…	
	TRUE	**FALSE**
Cannot reject H_0	Correct outcome	Type II error
Cannot accept H_0	Type I error	Correct outcome

Hypothesis Testing

Type I Error:
Rejecting the null hypothesis H_0 when it is true.

Type II Error:
Failing to reject the null hypothesis H_0 when it is false.

Each of the errors occurs with a particular probability:

α = probability of a Type I error = P(Type I error) = probability of rejecting the null hypothesis when the null hypothesis is true; rejecting a “good null”.

β = probability of a Type II error = P(Type II error) = probability of not rejecting the null hypothesis when the null hypothesis is false. $(1-\beta)$ is called the Power of the Test.

α and β should be as small as possible because they are probabilities of errors.

Hypothesis Testing (general guidelines)

1. Parameter of interest: From the problem context, identify the parameter of interest.
2. Null hypothesis, H_0: State the null hypothesis, H_0.
3. Alternative hypothesis, H_a: Specify an appropriate alternative hypothesis, H_a.
4. Test statistic: Determine an appropriate test statistic.
5. Reject H_0 if: State the rejection criteria for the null hypothesis.
6. Computations: Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
7. Draw conclusions: Decide whether or not H_0 should be rejected and report that in the problem context.

Steps 1−4 should be completed prior to examination of the sample data.