Recap

- What is type-1 vs type-2 error?
- What is the significance level α? What is the p-value?
- **Problem 1:** Consider a game of lotto ball. Many balls with numbers written on them are mixed in a big bowl. You are asked to bet on a number. Your reward is inversely proportional to the difference between your number and the number on the ball randomly pulled out of the big bowl. Which number will you bet on? Why?
- **Problem 2:** To test the newspaper claim that the mean wage rate of local foundry workers is > \$16 an hour, 25 foundry workers were randomly surveyed. It was found that the average wage rate for the sample of workers was \$14.50. Historical data suggest that the wage rates follow the normal distribution and the standard deviation of wage rates is \$3. Can the Union claim that the average wage *is less than \$16/hr*? Assume α =0.05.
- **Problem 3:** Who is a better player? Virat or Babar? (α = 5%)

Batsman	One sample each of 10 ODI innings	Sample Mean	Assume Same Population Std
Virat	00, 53, 34, 31, 00, 54, 96, 20, 10, 19	**31.7**	~~29.6~~ 28
Babar	12, 09, 91, 79, 51, 45, 41, 46, 29, 33	**43.6**	~~26.0~~ 28

Comparative Experiments

Case 3: What happens when we have NO population data?

Example 1

An engineer is studying **the formulation of a Portland cement mortar**. He has **added a polymer latex emulsion** during mixing to determine if this impacts the curing time and tension bond strength of the mortar.

The experimenter prepared 10 samples of the original formulation and 10 samples of the modified formulation.

Question: Does adding polymer latex emulsion change the strength?

■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Case 3: What happens when we have NO population data?

Example 2

Who is a better ODI batsman, Virat or Babar? (Based on the runs scored in an inning)

Batsman	One sample each of 10 ODI innings	Sample Mean	Sample Std. Dev
Virat	00, 53, 34, 31, 00, 54, 96, 20, 10, 19	**31.7**	29.6
Babar	12, 09, 91, 79, 51, 45, 41, 46, 29, 33	**43.6**	26.0

What is the hypothesis test?

What is the statistical (mathematical) model based on the hypothesis?

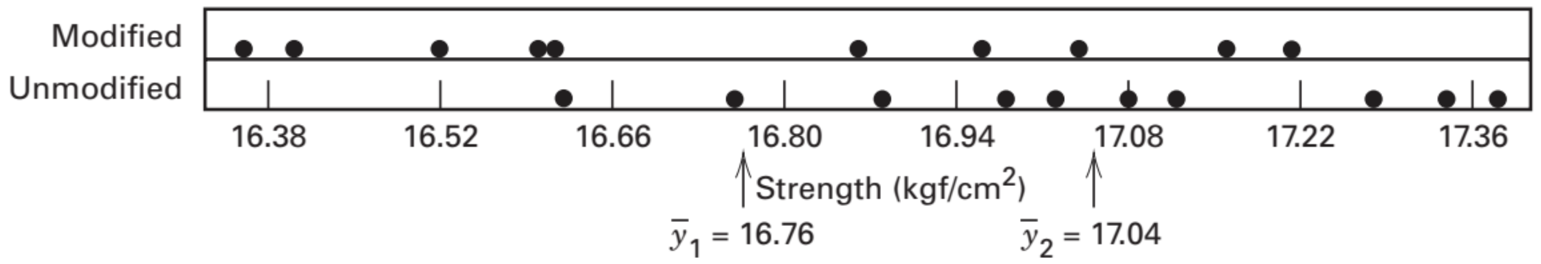
Example 1: Mortar Formula

■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Ref: Design and Analysis of Experiments, 8th Ed.

Each of the observations in the Portland cement experiment described above would be called a **run**. Notice that the individual runs differ, so there is fluctuation, or **noise**, in the observed bond strengths. This noise is usually called **experimental error** or simply **error**. It is a **statistical error**, meaning that it arises from variation that is uncontrolled and generally unavoidable. The presence of error or noise implies that the response variable, tension bond strength, is a **random variable**. A random variable may be either **discrete** or **continuous**. If the set of all possible values of the random variable is either finite or countably infinite, then the random variable is discrete, whereas if the set of all possible values of the random variable is an interval, then the random variable is continuous.



■ **FIGURE 2.1** **Dot diagram for the tension bond strength data in Table 2.1**

Example 1: Mortar Formula

Let $y_{11}, y_{12}, y_{13}, .. y_{1n1}$ be n_1 observations from the first factor level (Modified Mortar)

and $y_{21}, y_{22}, y_{23}, .. y_{2n1}$ be n_2 observations from the second factor level (UNmodified Mortar)

What is the hypothesis test?

A simple statistical model to describe the data is

$$y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \ldots, n_i \end{cases}$$

where y_{ij} is the jth observation from factor level i, μ_i is the mean of the response at the ith factor level, and ϵ_{ij} is a normal random variable associated with the ijth observation.

Ref: Design and Analysis of Experiments, 8th Ed.

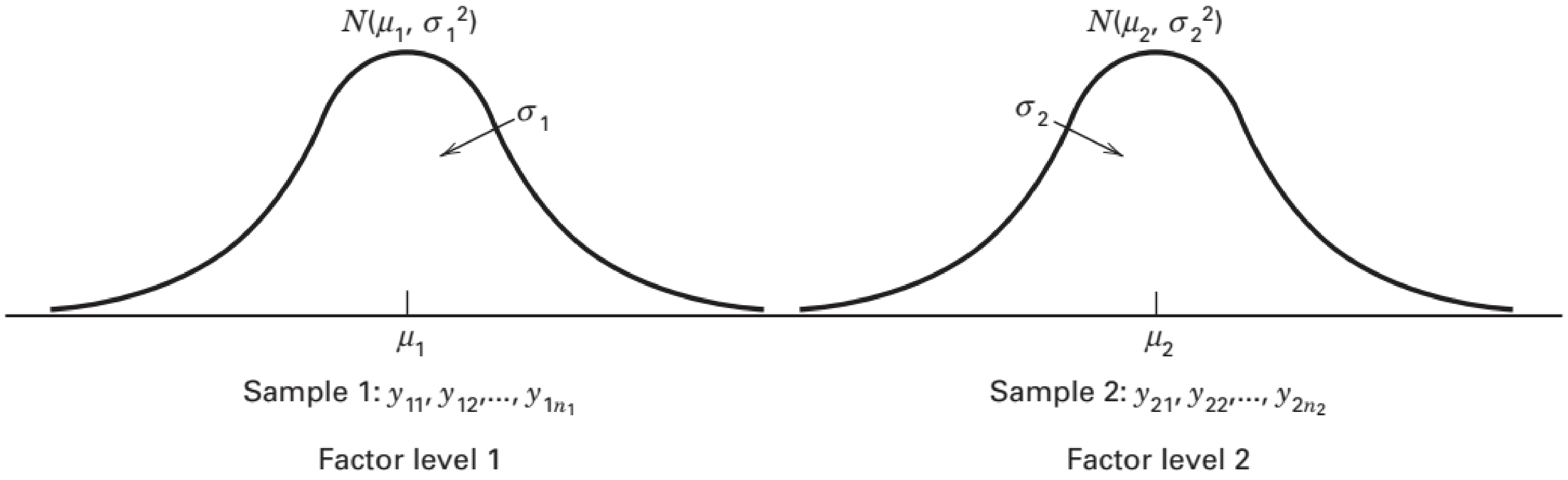
■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Example 1: Mortar Formula

We assume that the random error components ϵ_{1j} and ϵ_{2j} are normally distributed with means 0 and variances σ_1^2 and σ_2^2

Which would follow that the y_{1j} and y_{2j} are normally distributed with means μ_1 *and* μ_2 and variances σ_1^2 and σ_2^2



■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27