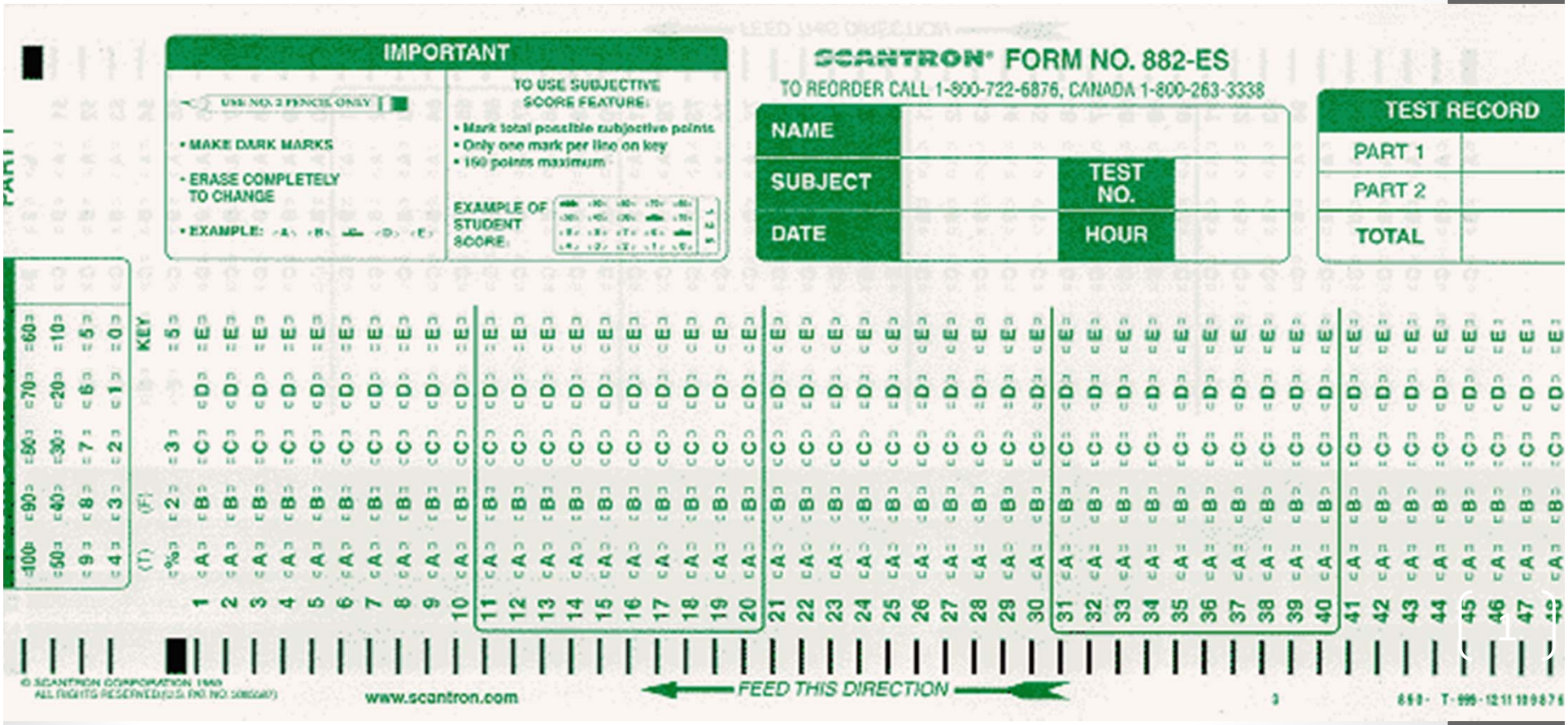
You will need eight (8) scantron forms 882-E



1.1 What is Statistics?

Definition. **Statistics** is a branch of mathematics that deals with the collection, organization, analysis, and interpretation of numerical observations.

Definition. Numerical observations in statistics are called **data**.

Note. The word "data" means "multiple observations". A single observation is called **datum** (rarely used).

Examples. It is correct to say "Data are obtained. They are analyzed. These data contain some variables. Data are messy. Collect data and analyze them. Data need to be analyzed."

Where are data collected?

- Stores – credit card use
- Doctor's office – medical records
- Payments of bills – cell phone, internet, water, gas, etc.
- Insurance companies – car, life, home insurances
- Banks – checking account, saving account, mortgage, student loan

- Sport events – football, basketball, baseball, Olympic games
- Social networks – facebook.com, linkedin.com
- Tax records
- Public records (birth, death, marriage)
- Census
- Etc.

1.2 Types of Statistics

There are two types of statistics: **descriptive** and **inferential.**

Definition. **Descriptive statistics** consists of methods of organizing, displaying, and describing data using tables, charts, graphs, and summary measures.

To define inferential statistics, we need to learn a few concepts first.

1.2 Population versus Sample

Definition. A **population** is a collection of all objects of interest.

Examples. All CSULB students, all CSULB freshmen, bikers in LA area.

Definition. A **sample** is a collection of several objects from a population.

Examples.

- students taking STAT 108 is a sample of all CSULB students
- students picked at different times and locations on campus is a sample of all CSULB students

What is the difference between these two samples (STAT 108 students versus students picked at different times and locations) ?

Definition. A sample is **random** if each object in the population is equally likely to be chosen for that sample. Otherwise, it is called **non-random**.

Examples.

- If students are picked at different times and locations on campus, then any CSULB student has an equal chance to be chosen, so it is a random sample.
- STAT 108 are only those CSULB students who are required to take this course, who have prerequisites, who haven't taken it yet, so it is a non-random sample.

Note. Intuitively, a random sample is preferred to a non-random sample because a random sample is more representative of the entire population.

Definition. **Inferential statistics** is a collection of methods that use results for a random sample to make decisions regarding the entire population.

Examples.

- The average GPA of a random sample of 100 students on campus may be generalized to all CSULB students.
- The average GPA in an honors class may not be generalized to all CSULB students because it is a non-random sample.

1.3 Basic Terms

Definition. Each object in a population or a sample is called a **member,** or an **observational unit**.

Definition. A **variable** is a characteristic under study that assumes different values for different observational units.

Examples. Age, gender, income, GPA, height, weight are variables.

Definition. A value that a variable assumes for a particular sample unit is called an **observation** (or **measurement**).

1.4 Types of Variables

There are two types of variables: **quantitative** (or **numerical**) and **qualitative** (or **categorical**).

Quantitative Variables

Definition. A **quantitative** (or **numerical**) variable is a variable that is measured numerically, and arithmetic operations on these numbers make sense.

Examples. Weight, height, income, age, number of pencils owned, length of daily commute.

Quantitative variables may be discrete or continuous.

Definition. A **discrete** variable assumes a finite or countably infinite values.

Examples. The number of cars in a parking lot, weight of a luggage in pounds, length of a box in inches, time spent watching TV in hours.

Definition. A **continuous** variable assumes values in an interval, that is, assumes non-countably infinite number of values.

Examples. Since weight, length, and time change continuously, they are naturally continuous variables, unless measured on some discrete scales (for example, pounds, inches, and hours).

Note. A discrete variable can be described as "the number of". For example,

- the number of cars in a parking lot
- the number of pounds a luggage weighs
- the number of inches the box measures
- the number of hours a person watches TV.

Qualitative Variables

Definition. A **qualitative** (or **categorical**) variable has values that fall into one of several categories. The values may be represented as numbers, but arithmetic operations would not make sense.

Examples. Gender (M/F), education (<HS, HSgrad, College+), course grade (A, B, C, D, or F).

Note. A quantitative variable may be turned into qualitative by introducing several categories.

Example. The time spent waiting for a bus is a quantitative variable (continuous, or discrete if measured in minutes).

It can be made qualitative by placing the wait time in one of, say, three categories: < 2 minutes, 2-5 minutes, and 5 or more minutes.