

Statistical Concepts For Your First Test

0. Business Statistics (Italian origin, relevant to State) provides justifiable answers to the following concerns for every consumer and producer (and service provider):

- What is your or your customer's, **Expectation** of the product/service you sell or that your customer buys? That is, what is a good estimate for μ ?
- Given the information about your or your customer's, expectation, what is the **Quality** of the product/service you sell or that your customers buy from your products or services? That is, what is a good estimate for quality (e.g., σ , or **C.V.= $100\sigma/\mu\%$**)?
- Given the information about your or your customer's expectation, and the quality of the product/service you sell or you customer buy, how does the product/service **compare** with other existing similar types? That is, comparing several μ 's, and several σ 's.
- Finding any (linear) **relationship for prediction** purposes. For example, sales (S) as function of advertising rate (A) for a specific budget, for a duration of campaign (T). $S = mA + b$.

1. Statistical techniques are methods that convert data (information) into useful knowledge to make **decision under uncertainty**. **Descriptive** techniques describe and summarize data to extract useful information. **Inferential** techniques allow us to make **estimates** and draw conclusions about **population** from samples. We need a large number of statistical techniques because there are numerous **statistical objectives**.

2. There different classifications of data: **quantitative** (real numbers) vs **qualitative** (i.e., categories), **continuous** (measurable) vs **discrete** (countable), **NOIR** (nominal, ordinal, interval, and ratio). Each type of data and statistical objective requires specific techniques.

3. We gather data by various **sampling** plans. However, the validity of any statistical outcome depends on the **validity** of the sampling, such as **randomness**.

4. The **sampling distribution** is the source of statistical inference. However, the interval estimator and the test statistic are derived by extending sampling information to population with certain level of uncertainty expressed by **probability** (to probe for something that is difficult to know). Examples of probabilities are **significant levels**, and **confidence levels**. There is a **duality between confidence interval and test of hypothesis**.

5. All inferences are actually probability statements based on **extending our knowledge from the sample to the population**. Because the **probability of an event is defined as the proportion of times the event occurs** in the long run, we must **interpret confidence interval** estimates in these terms.

6. All tests of hypotheses are conducted similarly. We need to set up the **null hypothesis (=)**, **alternative hypothesis**, and a desirable confidence level that determines the sample size n . We then compute the value of the **test statistic**. If the difference between what we have **observed** (i.e. calculated) and what we **expect to observe** (the **critical values** for a related **statistical table**) is too large, we reject the null hypothesis. The standard that decides what is "too large" is determined by the probability of a **Type I error** and the **sample size (of at least 2, d.f. $n - 1$)**.

7. In any test of hypothesis (and in most decisions) there are two possible errors, **Type I and Type II errors**. The relationship between these errors helps us decide where to set the standard. If we set the standard so high that the probability of a Type I error is very small, we increase the probability of a Type II error. A procedure designed to decrease the probability of a Type II error

must have a relatively large probability of a Type I error. To reduce both errors one must have more information, i.e., larger sample size(s).

8. When you take a sample from a population and compute the sample mean clearly different samples (almost always) result in different means, thus it is **random variable** having a **density function**. The distribution of all possible sample means values, for sample of a particular fixed size, is called the **sampling distribution (of means)**.

9. The variability of the distribution of sample means depends on how large your sample is and on how much variability there is in the population from which the samples are taken. As the **size of the sample** increases, **the variability of the sample means decreases** (for variance, by factor of $1/n$).

10. A **normal distribution** is bell-shaped. It is a symmetric distribution in which **the mean, median, and mode** all coincide. Many natural random variables, such as height and weight, have distributions that are approximately normal. The mean of **standard normal** is 0, and the standard deviation is 1. Using **Z score** all normal distributions will be reduced to the standard normal.

11. The **Central Limit Theorem** states that for samples of a **sufficiently large (fixed) size**, the distribution of sample means is approximately normal. That's why the normal distribution is so important in **data analysis for population mean** μ : Testing and confidence interval
<http://home.ubalt.edu/ntsbarsh/Business-stat/StatSummaySheet.pdf>

12. A **confidence interval** provides a range of values that, with a designated **likelihood (i.e. probability, chance, odds)**, contains e.g. the population mean μ , or standard deviation σ .

13. The variability in a population can be estimated by sample **standard deviation**. It is an indication of **"good" quality** if the values of random are close to, what one expects, i.e. the expected value. The **variability is related to quality** because one of the key characteristics of quality is consistency, i.e., having low variation; that is needed for good managerial planning.

14. All statistical studies required that the **population be homogeneous** and the **sample be random**. The first requirement is checked by **relative (frequency) histogram** must have one single mode; the **randomness is checked by Runs-test**.

15. The statistical tables are used to find **critical values** are the standard normal, Student-t, χ^2 and the F table (The **last three required the Normality** of the populations).

16. What **statistical table** to use? <http://home.ubalt.edu/ntsbarsh/Business-stat/selection.gif>

17. **Statistical skills** enable us to intelligently collect, analyze and interpret data relevant to their **decision-making**. **Statistical concepts** and **statistical thinking** enable them to: **Solve problems in a diversity of contexts, Add substance to decisions, Reduce guesswork.**