

# CHAPTER 18:

## Inference in Practice

**Basic Practice of  
Statistics**

7<sup>th</sup> Edition

**Lecture PowerPoint Slides**

# In Chapter 18, We Cover ...

- Conditions for inference in practice
- Cautions about confidence intervals
- Cautions about significance tests
- Planning studies: sample size for confidence intervals
- Planning studies: the power of a statistical test\*

# z Procedures

- So far, we have met two procedures for statistical inference. When the “simple conditions” are true—the data are an SRS, the population has a Normal distribution, and we know the standard deviation  $\sigma$  of the population—a confidence interval for the mean  $\mu$  is

$$\bar{x} \pm z^* \sigma / \sqrt{n}$$

- To test a hypothesis  $H_0: \mu = \mu_0$ , we use the one-sample  $z$ -statistic:

$$\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

- These are called **z procedures** because they both involve a one-sample  $z$ -statistic and use the standard Normal distribution.

# Conditions for Inference in Practice

- Any confidence interval or significance test can be trusted only under specific conditions.

## **WHERE THE DATA COME FROM MATTERS**

- When you use statistical inference, you are acting as if your data are a random sample or come from a randomized comparative experiment.
- If your data don't come from a random sample or randomized comparative experiment, your conclusions may be challenged.
- Practical problems such as nonresponse or dropouts from an experiment can hinder inference.
- Different methods are needed for different designs.
- There is no cure for fundamental flaws like voluntary response.

# Conditions for Inference in Practice

- **What is the shape of the population distribution?**
- Many of the basic methods of inference are designed for Normal populations.
- Fortunately, this condition is less essential than where the data come from.
- Any inference procedure based on sample statistics (like the sample mean,  $\bar{x}$ ) that are not resistant to outliers can be strongly influenced by a few extreme observations.

# Cautions About Confidence Intervals

- A sampling distribution shows how a statistic varies in repeated random sampling.
- This variation causes random sampling error because the statistic misses the true parameter by a random amount.
- So the margin of error in a confidence interval ignores everything except the sample-to-sample variation due to choosing the sample randomly.

## **THE MARGIN OF ERROR DOESN'T COVER ALL ERRORS**

- The margin of error in a confidence interval covers only random sampling errors.
- Practical difficulties such as undercoverage and nonresponse are often more serious than random sampling error. The margin of error does not take such difficulties into account.

# Cautions About Significance Tests

- Significance tests are widely used in most areas of statistical work. Some points to keep in mind when you use or interpret significance tests are:
- ***How small a  $P$  is convincing?***
- The purpose of a test of significance is to describe the degree of evidence provided by the sample against the null hypothesis. How small a  $P$ -value is convincing evidence against the null hypothesis depends mainly on two circumstances:
  - If  $H_0$  represents an assumption that has been believed for years, strong evidence (a small  $P$ ) will be needed.
  - If rejecting  $H_0$  means making a costly changeover, you need strong evidence.

# Cautions About Significance Tests

- ***Significance depends on the alternative hypothesis.***
- The  $P$ -value for a one-sided test is one-half the  $P$ -value for the two-sided test of the same null hypothesis based on the same data.
  - The evidence against the null hypothesis is stronger when the alternative is one-sided because it is based on the data *plus* information about the direction of possible deviations from the null.
  - If you lack this added information, always use a two-sided alternative hypothesis.



# Cautions About Significance Tests

## **SAMPLE SIZE AFFECTS STATISTICAL SIGNIFICANCE**

- Because large random samples have small chance variation, very small population effects can be highly significant if the sample is large.
- Because small random samples have a lot of chance variation, even large population effects can fail to be significant if the sample is small.
- Statistical significance does not tell us whether an effect is large enough to be important. That is, **statistical significance is not the same as practical significance.**

## ***Beware of Multiple Analyses***

- The reasoning of statistical significance works well if you decide what effect you are seeking, design a study to search for it, and use a test of significance to weigh the evidence you get.

# Planning Studies: Sample Size for Confidence Intervals

- A wise user of statistics never plans a sample or an experiment without also planning the inference. The number of observations is a critical part of planning the study.
- The margin of error ME of the confidence interval for the population mean  $\mu$  is

$$m = z^* \frac{\sigma}{\sqrt{n}}$$

- To obtain a desired margin of error  $m$ , put in the value of  $z^*$  for your desired confidence level, and solve for the sample size  $n$ .

## **SAMPLE SIZE FOR DESIRED MARGIN OF ERROR**

- The  $z$  confidence interval for the mean of a Normal population will have a specified margin of error  $m$  when the sample size is

$$n = \left( \frac{z^* \sigma}{m} \right)^2$$

# Planning Studies: the Power of a Statistical Test\*

- How large a sample should we take when we plan to carry out a test of significance?
- Here are the questions we must answer to decide how many observations we need:
  - **Significance level.** How much protection do we want against getting a significant result from our sample when there really is no effect in the population?
  - **Effect size.** How large an effect in the population is important in practice?
  - **Power.** How confident do we want to be that our study will detect an effect of the size we think is important?

# Planning Studies: the Power of a Statistical Test\*

- The **power** of a test against a specific alternative is the probability that the test will reject  $H_0$  at a chosen significance level  $\alpha$  when the specified alternative value of the parameter is true.
- Example: a matched-pairs experiment to test for loss of sweetness—the bigger the difference, the bigger the loss of sweetness. We know the sweetness loss scores vary from scorer to scorer according to a Normal distribution with standard deviation about  $\sigma = 1$ . To see if there is a loss in sweetness, we test

$$H_0: \mu = 0$$

$$H_a: \mu > 0$$

- Will 10 tasters be sufficient for this study?

# Planning Studies: the Power of a Statistical Test\*

- Example (cont'd): Will 10 tasters be sufficient for this study, or should we use more?
- **Significance level.** Requiring significance at the 5% level is enough protection against declaring there is a loss in sweetness when in fact there is no change if we could look at the entire population.
- **Effect size.** A mean sweetness loss of 0.8 point on the 10-point scale will be noticed by consumers and so is important in practice.
- **Power.** We want to be 90% confident that our test will detect a mean loss of 0.8 point in the population of all tasters.

Difference	Sample Size	Target Power	Actual Power
0.1	857	0.9	0.900184
0.2	215	0.9	0.901079
0.3	96	0.9	0.902259
0.4	54	0.9	0.902259
0.5	35	0.9	0.905440
0.6	24	0.9	0.902259
0.7	18	0.9	0.907414
0.8	14	0.9	0.911247
0.9	11	0.9	0.909895
1.0	9	0.9	0.912315

- Result (Minitab output, left): We see that our earlier sample of 10 tasters is not large enough to be 90% confident of detecting (at the 5% significance level) an effect of size 0.8. If we want power 90% against effect size 0.8, we need at least 14 tasters. The actual power with 14 tasters is 0.911247.

# Planning Studies: the Power of a Statistical Test\*

- How large a sample should we take when we plan to carry out a significance test? The answer depends on what alternative values of the parameter are important to detect.
- Here is an overview of influences on “How many observations do I need?”
- If you insist on a smaller significance level (such as 1% rather than 5%), you have to take a larger sample. A smaller significance level requires stronger evidence to reject the null hypothesis.
- If you insist on higher power (such as 99% rather than 90%), you will need a larger sample. Higher power gives a better chance of detecting a difference when it is really there.
- At any significance level and desired power, a two-sided alternative requires a larger sample than a one-sided alternative.
- At any significance level and desired power, detecting a small effect requires a larger sample than detecting a large effect.