Chapters 14 and 15

Introduction to Inference

Statistical Inference

- Provides methods for drawing conclusions about a population from sample data
  - Confidence Intervals
    - What is the population mean?
  - Tests of Significance
    - Is the population mean larger than 66.5?

Inference about a Mean

Simple Conditions

1. SRS from the population of interest
2. Variable has a Normal distribution $N(\mu, \sigma)$ in the population
3. Although the value of μ is unknown, the value of the population standard deviation σ is known

Point Estimator

By the Law of Large Numbers:
The sample mean is a good estimate of the true mean.

The sample mean is a "point estimator."
It estimates the value of the parameter.

But how "confident" are we?

Confidence Interval

A ***level C confidence interval*** has two parts

1. An interval calculated from the data, usually of the form:
   estimate ± margin of error
2. The ***confidence level C***, which is the probability that the interval will capture the true parameter value in repeated samples; that is, ***C*** is the success rate for the method. Measures how "confident" we are.

Case Study

NAEP Quantitative Scores
(National Assessment of Educational Progress)

Rivera-Batiz, F. L., "Quantitative literacy and the likelihood of employment among young adults," *Journal of Human Resources*, **27** (1992), pp. 313-328.

What is the average score for all young adult males?

Case Study

NAEP Quantitative Scores

The NAEP survey includes a short test of quantitative skills, covering mainly basic arithmetic and the ability to apply it to realistic problems. Scores on the test range from 0 to 500, with higher scores indicating greater numerical abilities. It is known that NAEP scores have standard deviation σ = 60.

BPS - 5th Ed. Chapter 14 7

Case Study

NAEP Quantitative Scores

In a recent year, 840 men 21 to 25 years of age were in the NAEP sample. Their mean quantitative score was 272.

On the basis of this sample, estimate the mean score μ in the population of all 9.5 million young men of these ages.

BPS - 5th Ed. Chapter 14 8

Case Study

NAEP Quantitative Scores

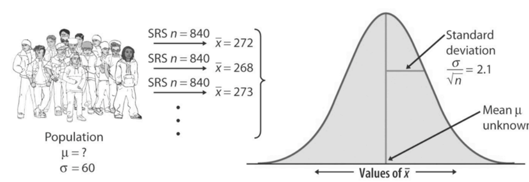
1. To estimate the unknown population mean μ, use the sample mean $\bar{x}$ = 272.
2. The law of large numbers suggests that $\bar{x}$ will be close to μ, but there will be some error in the estimate.
3. The sampling distribution of $\bar{x}$ has the Normal distribution with mean μ and standard deviation

$$\frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{840}} \approx 2.1$$

BPS - 5th Ed. Chapter 14 9

Case Study

NAEP Quantitative Scores

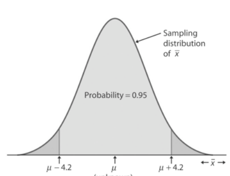


BPS - 5th Ed. Chapter 14 10

Case Study

NAEP Quantitative Scores

4. The 68-95-99.7 rule indicates that $\bar{x}$ and μ are within two standard deviations (4.2) of each other in about 95% of all samples.



$$\bar{x} - 4.2 = 272 - 4.2 = 267.8$$
$$\bar{x} + 4.2 = 272 + 4.2 = 276.2$$

BPS - 5th Ed. Chapter 14 11

Case Study

NAEP Quantitative Scores

So, if we estimate that μ lies within 4.2 of $\bar{x}$, we'll be right about 95% of the time.

BPS - 5th Ed. Chapter 14 12

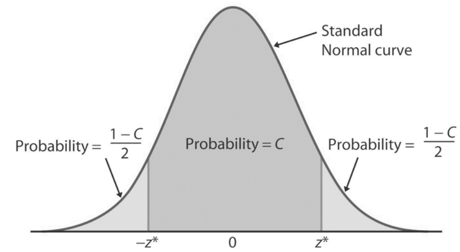
Confidence Interval
Mean of a Normal Population

Take an SRS of size *n* from a Normal population with unknown mean μ and known standard deviation σ. A level C confidence interval for μ is:

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

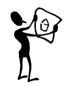
BPS - 5th Ed. Chapter 14 13

Confidence Interval
Mean of a Normal Population



BPS - 5th Ed. Chapter 14 14

Case Study
NAEP Quantitative Scores



Using the 68-95-99.7 rule gave an approximate 95% confidence interval. A more precise 95% confidence interval can be found using the appropriate value of *z** (1.960) with the previous formula.

$$\bar{x} - (1.960)(2.1) = 272 - 4.116 = 267.884$$

$$\bar{x} + (1.960)(2.1) = 272 + 4.116 = 276.116$$

We are 95% confident that the average NAEP quantitative score for all adult males is between 267.884 and 276.116.

BPS - 5th Ed. Chapter 14 15

Careful Interpretation of a Confidence Interval

- ***"We are 95% confident that the mean NAEP score for the population of all adult males is between 267.884 and 276.116."***
  (We feel that plausible values for the population of males' mean NAEP score are between 267.884 and 276.116.)
- ** This *does not mean* that 95% of all males will have NAEP scores between 267.884 and 276.116. **
- *Statistically*: 95% of the C.I.s should contain the true population mean.
- Again: 5% of the polls you've heard so far were wrong!

BPS - 5th Ed. Chapter 14 16

How Confidence Intervals Behave

- The margin of error is:

$$\text{margin of error} = z^* \frac{\sigma}{\sqrt{n}}$$

- The margin of error gets smaller, resulting in more accurate inference,
  - when *n* gets larger
  - when *z** gets smaller (confidence level gets smaller)
  - when σ gets smaller (less variation)

BPS - 5th Ed. Chapter 15 17

Case Study

NAEP Quantitative Scores (Ch. 14)

95% Confidence Interval

$$\bar{x} - (1.960)(2.1) = 272 - 4.116 = 267.884$$

$$\bar{x} + (1.960)(2.1) = 272 + 4.116 = 276.116$$

90% Confidence Interval

$$\bar{x} - (1.645)(2.1) = 272 - 3.4545 = 268.5455$$

$$\bar{x} + (1.645)(2.1) = 272 + 3.4545 = 275.4545$$

The 90% CI is narrower than the 95% CI.

BPS - 5th Ed. Chapter 15 18

Planning Studies

Choosing the Sample Size for a C.I.

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

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

Case Study

NAEP Quantitative Scores (Ch.14)

Suppose that we want to estimate the population mean NAEP scores using a 90% confidence interval, and we are instructed to do so such that the margin of error does not exceed 3 points (recall that $\sigma = 60$).

What sample size will be required to enable us to create such an interval?

Case Study

NAEP Quantitative Scores

$$n = \left(\frac{z^{*}\sigma}{m} \right)^{2} = \left(\frac{(1.645)(60)}{3} \right)^{2} = 1082.41$$

Thus, we will need to sample at least 1082.41 men aged 21 to 25 years to ensure a margin of error not to exceed 3 points.
Note that since we can't sample a fraction of an individual and using 1082 men will yield a margin of error slightly more than 3 points, our sample size should be $n = 1083$ men.

Cautions About Confidence Intervals

The margin of error does not cover all errors.

- The margin of error in a confidence interval covers only random sampling errors. No other source of variation or bias in the sample data influence the sampling distribution.
- 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.

Be aware of these points when reading any study results.

What are Tests of Significance?

- Claim: "John gets 80% of free shots"
- Data: results on 1000 free shots.
- Law of large numbers: average number of shots scored should be close to the true scoring percentage.
- Data: average number of shots scored is 60%.

Reasoning of Tests of Significance

- How likely would it be to see the results we saw if the claim were true?
  (if John truly gets 80% of his shots, how likely is he to get only 60% in 1000 shots?!!)
- Do the data give enough evidence against the claim?
  (what if John scored 75%? 20%?)

Stating Hypotheses
Null Hypothesis, H_0

- The statement being tested in a statistical test is called the **null hypothesis**.
- The test is designed to assess the strength of evidence **against** the null hypothesis.
- Usually the null hypothesis is a statement of “no effect” or “no difference”, or it is a statement of equality.
- **When performing a hypothesis test, we assume that the null hypothesis is true until we have sufficient evidence against it.**

Stating Hypotheses
Alternative Hypothesis, H_a

- The statement we are trying to find evidence for is called the **alternative hypothesis**.
- Usually the alternative hypothesis is a statement of “there is an effect” or “there is a difference”, or it is a statement of inequality.
- **The alternative hypothesis should express the hopes or suspicions we bring to the data. <u>It is cheating to first look at the data and then frame H_a to fit what the data show</u>.**

Case Study I
Sweetening Colas

Diet colas use artificial sweeteners to avoid sugar. These sweeteners gradually lose their sweetness over time. Trained testers sip the cola and assign a “sweetness score” of 1 to 10. The cola is then retested after some time and the two scores are compared to determine the difference in sweetness after storage. Bigger differences indicate bigger loss of sweetness.

Case Study I
Sweetening Colas

Suppose we know that for any cola, the sweetness loss scores vary from taster to taster according to a Normal distribution with standard deviation $\sigma = 1$.

The mean μ for all tasters measures loss of sweetness.

The sweetness losses for a new cola, as measured by 10 trained testers, yields an average sweetness loss of $\bar{x} = 1.02$. **Do the data provide sufficient evidence that the new cola lost sweetness in storage?**

Case Study I
Sweetening Colas

The null hypothesis is no average sweetness loss occurs, while the alternative hypothesis (that which we want to show is likely to be true) is that an average sweetness loss does occur.

H_0: $\mu = 0$ $\quad\quad$ H_a: $\mu > 0$

This is considered a one-sided test because we are interested only in determining if the cola lost sweetness (gaining sweetness is of no consequence in this study).

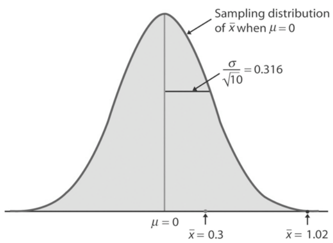
Case Study I
Sweetening Colas

- If the claim that $\mu = 0$ is true (no loss of sweetness, on average), the sampling distribution of $\bar{x}$ from 10 tasters is Normal with $\mu = 0$ and standard deviation

$$\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{10}} \approx 0.316$$

- The data yielded $\bar{x} = 1.02$, which is more than three standard deviations from $\mu = 0$. This is strong evidence that the new cola lost sweetness in storage.
- If the data yielded $\bar{x} = 0.3$, which is less than one standard deviations from $\mu = 0$, there would be no evidence that the new cola lost sweetness in storage.

Case Study I
Sweetening Colas



The Hypotheses for Means

- Null: H_0: $\mu = \mu_0$
- One sided alternatives

  H_a: $\mu > \mu_0$

  H_a: $\mu < \mu_0$
- Two sided alternative

  H_a: $\mu \neq \mu_0$

Case Study II
Studying Job Satisfaction

Does the job satisfaction of assembly workers differ when their work is machine-paced rather than self-paced? A matched pairs study was performed on a sample of workers, and each worker's satisfaction was assessed after working in each setting. The response variable is the difference in satisfaction scores, self-paced minus machine-paced.

Case Study II
Studying Job Satisfaction

The null hypothesis is no average difference in scores in the population of assembly workers, while the alternative hypothesis (that which we want to show is likely to be true) is there is an average difference in scores in the population of assembly workers.

H_0: $\mu = 0$ $\qquad$ H_a: $\mu \neq 0$

This is considered a two-sided test because we are interested determining if a difference exists (the direction of the difference is not of interest in this study).

Test Statistic
Testing the Mean of a Normal Population

Take an SRS of size *n* from a Normal population with unknown mean μ and known standard deviation σ. The test statistic for hypotheses about the mean (H_0: $\mu = \mu_0$) of a Normal distribution is the standardized version of $\bar{x}$:

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

(called the z-score)

Case Study I
Sweetening Colas

If the null hypothesis of no average sweetness loss is true, the test statistic would be:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{1.02 - 0}{1 / \sqrt{10}} \approx 3.23$$

Because the sample result is more than 3 standard deviations above the hypothesized mean 0, it gives strong evidence that the mean sweetness loss is not 0, but positive.

P-value

Assuming that the null hypothesis is true, the probability that the test statistic would take a value as extreme or more extreme than the value actually observed is called the ***P*-value** of the test.

The smaller the *P*-value, the stronger the evidence the data provide against the null hypothesis. That is, a small *P*-value indicates a small likelihood of observing the sampled results if the null hypothesis were true.

BPS - 5th Ed. Chapter 14 37

P-value for Testing Means

- H_a: $\mu > \mu_0$
  - *P*-value is the probability of getting a value as large or larger than the observed z-score.
- H_a: $\mu < \mu_0$
  - *P*-value is the probability of getting a value as small or smaller than the observed z-score.
- H_a: $\mu \neq \mu_0$
  - *P*-value is *two times* the probability of getting a value as large or larger than the absolute value of the observed z-score.

BPS - 5th Ed. Chapter 14 38

Case Study I

Sweetening Colas

For test statistic $z = 3.23$ and alternative hypothesis H_a: $\mu > 0$, the *P*-value would be:

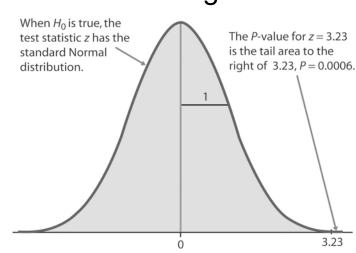
$$P\text{-value} = P(Z > 3.23) = 1 - 0.9994 = 0.0006$$

If H_0 is true, there is only a 0.0006 (0.06%) chance that we would see results at least as extreme as those in the sample; thus, since we saw results that are unlikely if H_0 is true, we therefore have evidence against H_0 and in favor of H_a.

BPS - 5th Ed. Chapter 14 39

Case Study I

Sweetening Colas



BPS - 5th Ed. Chapter 14 40

Case Study II

Studying Job Satisfaction

Suppose job satisfaction scores follow a Normal distribution with standard deviation $\sigma = 60$. Data from 18 workers gave a sample mean score of 17. If the null hypothesis of no average difference in job satisfaction is true, the test statistic would be:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{17 - 0}{60/\sqrt{18}} \approx 1.20$$

BPS - 5th Ed. Chapter 14 41

Case Study II

Studying Job Satisfaction

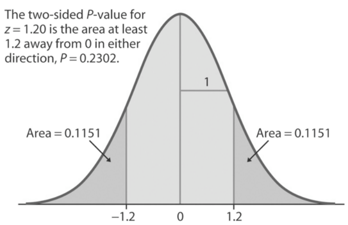
For test statistic $z = 1.20$ and alternative hypothesis H_a: $\mu \neq 0$, the *P*-value would be:

$$\begin{aligned} P\text{-value} &= P(Z < -1.20 \text{ or } Z > 1.20) \\ &= 2\,P(Z < -1.20) = 2\,P(Z > 1.20) \\ &= (2)(0.1151) = 0.2302 \end{aligned}$$

If H_0 is true, there is a 0.2302 (23.02%) chance that we would see results at least as extreme as those in the sample; thus, since we saw results that are likely if H_0 is true, we therefore do not have good evidence against H_0 and in favor of H_a.

BPS - 5th Ed. Chapter 14 42

Case Study II
Studying Job Satisfaction



BPS - 5th Ed. Chapter 14 43

Statistical Significance

- Tells how small is small!
- If the *P*-value is as small as or smaller than the **significance level α** (i.e., *P*-value ≤ α), then we CONCLUDE that data give results that are **statistically significant at level α**.
- If we choose α = 0.05, we are requiring that the data give evidence against H_0 so strong that it would occur no more than 5% of the time when H_0 is true.
- If we choose α = 0.01, we are insisting on stronger evidence against H_0, evidence so strong that it would occur only 1% of the time when H_0 is true. (We are more strict)

BPS - 5th Ed. Chapter 14 44

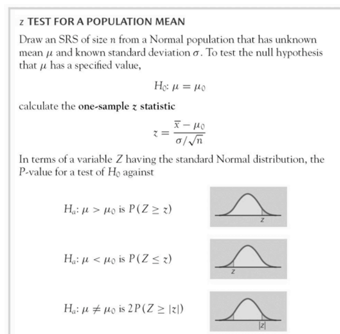
Tests for a Population Mean

The four steps in carrying out a significance test:

1. State the null and alternative hypotheses.
2. Calculate the test statistic.
3. Find the *P*-value.
4. State your **CONCLUSION WITHOUT MATH.**

The procedure for Steps 2 and 3 is on the next page.

BPS - 5th Ed. Chapter 14 45



***z* TEST FOR A POPULATION MEAN**

Draw an SRS of size n from a Normal population that has unknown mean μ and known standard deviation σ. To test the null hypothesis that μ has a specified value,

$$H_0: \mu = \mu_0$$

calculate the **one-sample z statistic**

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

In terms of a variable Z having the standard Normal distribution, the P-value for a test of H_0 against

$H_a: \mu > \mu_0$ is $P(Z \geq z)$

$H_a: \mu < \mu_0$ is $P(Z \leq z)$

$H_a: \mu \neq \mu_0$ is $2P(Z \geq |z|)$

BPS - 5th Ed. Chapter 14 46

Case Study I
Sweetening Colas

1. ***Hypotheses:*** H_0: $\mu = 0$
   H_a: $\mu > 0$
2. ***Test Statistic:***
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.02 - 0}{1/\sqrt{10}} \approx 3.23$$
3. ***P-value:*** *P*-value = $P(Z > 3.23) = 1 - 0.9994 = 0.0006$
4. ***CONCLUSION:***
   Since the *P*-value is smaller than α = 0.01, **there is very strong evidence that the new cola loses sweetness on average during storage at room temperature.**

BPS - 5th Ed. Chapter 14 47

Case Study II
Studying Job Satisfaction

1. ***Hypotheses:*** H_0: $\mu = 0$
   H_a: $\mu \neq 0$
2. ***Test Statistic:***
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{17 - 0}{60/\sqrt{18}} \approx 1.20$$
3. ***P-value:*** *P*-value = $2P(Z > 1.20) = (2)(1 - 0.8849) = 0.2302$
4. ***CONCLUSION:***
   Since the *P*-value is larger than α = 0.10, **there is not sufficient evidence that mean job satisfaction of assembly workers differs when their work is machine-paced rather than self-paced.**

BPS - 5th Ed. Chapter 14 48

Confidence Intervals & Two-Sided Tests

A level α two-sided significance test rejects the null hypothesis H_0: $\mu = \mu_0$ exactly when the value μ_0 falls outside a level $(1 - \alpha)$ confidence interval for μ.

Case Study II
Studying Job Satisfaction

A 90% confidence interval for μ is:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 17 \pm 1.645 \frac{60}{\sqrt{18}} = 17 \pm 23.26$$

$$= -6.26 \text{ to } 40.26$$

Since $\mu_0 = 0$ is in this confidence interval, it is plausible that the true value of μ is 0; thus, there is not sufficient evidence (at $\alpha = 0.10$) that the mean job satisfaction of assembly workers differs when their work is machine-paced rather than self-paced.

z Procedures

- If we know the standard deviation σ of the population, a confidence interval for the mean μ is:

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

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

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

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

Conditions for Inference in Practice

- The data must be an SRS from the population (ask: "where did the data come from?").
  - Different methods are needed for different designs.
  - The *z* procedures are not correct for samples other than SRS.
- Outliers can distort the result.
  - The sample mean is strongly influenced by outliers.
  - Always explore your data before performing an analysis. (But design the test FIRST!! Otherwise, data is corrupted!)
- The shape of the population distribution matters.
  - Skewness and outliers make the *z* procedures untrustworthy unless the sample is large.
  - In practice, the *z* procedures are reasonably accurate for samples of at least moderate size from a fairly symmetric distribution.
- The population standard deviation σ must be known.
  - Unfortunately σ is rarely known, so *z* procedures are rarely useful.
  - Chapter 17 will introduce procedures for when σ is unknown.

Where Did the Data Come From?

- When you use statistical inference, you are acting as if your data are a probability sample or come from a randomized experiment.
- Statistical confidence intervals and tests cannot remedy basic flaws in producing data, such as voluntary response samples or uncontrolled experiments. Also be aware of nonresponse or dropouts in well-designed studies.
- If the data do not come from a probability sample or a randomized experiment, the conclusions may be open to challenge. To answer the challenge, ask whether the data can be trusted as a basis for the conclusions of the study.

Case Study
Mammary Artery Ligation

Barsamian, E. M., "The rise and fall of internal mammary artery ligation," *Costs, Risks, and Benefits of Surgery*, Bunker, Barnes, and Mosteller (eds.), Oxford University Press, 1977, pp. 212-220.

Surgeons tested a procedure to alleviate pain caused by inadequate blood supply to the heart, and the patients reported a statistically significant reduction in angina pain.

Case Study
Mammary Artery Ligation

Statistical significance indicates that something other than chance is at work, but it does not say what that something is. Since this experiment was not controlled, the reduction in pain could be due to the placebo effect. A controlled experiment showed that this was the case, and surgeons immediately stopped performing the operation.

BPS - 5th Ed. Chapter 15 55

Cautions About Significance Tests
How small a P-value is convincing?

- If H_0 represents an assumption that people have believed in for years, strong evidence (small *P*-value) will be needed to persuade them otherwise.
- If the consequences of rejecting H_0 are great (such as making an expensive or difficult change from one procedure or type of product to another), then strong evidence as to the benefits of the change will be required.

Although α = 0.05 is a common cut-off for the P-value, there is no set border between "significant" and "insignificant," only increasingly strong evidence against H_0 (in favor of H_a) as the P-value gets smaller.

BPS - 5th Ed. Chapter 15 56

Cautions About Significance Tests
Significance depends on the Alternative Hyp.

- 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 H_0 is stronger when the alternative is one-sided; use one-sided tests if you know the direction of possible deviations from H_0, otherwise you must use a two-sided alternative.

BPS - 5th Ed. Chapter 15 57

Cautions About Significance Tests
Statistical Significance & Practical Significance (and the effect of Sample Size)

- When the sample size is very large, tiny deviations from the null hypothesis (with little practical consequence) will be statistically significant.
- When the sample size is very small, large deviations from the null hypothesis (of great practical importance) might go undetected (statistically insignificant).

Statistical significance is not the same thing as practical significance.

BPS - 5th Ed. Chapter 15 58

Case Study: Drug Use in American High Schools
Alcohol Use

Bogert, Carroll. "Good news on drugs from the inner city," *Newsweek*, Feb.. 1995, pp 28-29.

BPS - 5th Ed. Chapter 15 59

Case Study
Alcohol Use

- Alternative Hypothesis: The percentage of high school students who used alcohol in 1993 is less than the percentage who used alcohol in 1992.
- Null Hypothesis: There is no difference in the percentage of high school students who used in 1993 and in 1992.

BPS - 5th Ed. Chapter 15 60

Case Study
Alcohol Use

1993 survey was based on 17,000 seniors, 15,500 10th graders, and 18,500 8th graders.

Grade	1992	1993	Diff	*P*-value
8th	53.7	51.6	-2.1	<.001
10th	70.2	69.3	-0.9	.04
12th	76.8	76.0	-0.8	.04

BPS - 5th Ed. Chapter 15 61

Case Study
Alcohol Use

- The article suggests that the survey reveals "good news" since the differences are all negative.
- The differences are *statistically significant*.
  - All *P*-values are less than $\alpha = 0.05$.
- The 10th and 12th grade differences probably are not *practically significant*.
  - Each difference is less than 1%

BPS - 5th Ed. Chapter 15 62

Case Study: Memory Loss in American Hearing, American Deaf, and Chinese Adults
Memory Loss

Levy, B. and E. Langer. "Aging free from negative stereotypes: Successful memory in China and among the American deaf," *Journal of Personality and Social Psychology*, Vol. 66, pp 989-997.

BPS - 5th Ed. Chapter 15 63

Case Study
Memory Loss

- Average Memory Test Scores (higher is better)
- 30 subjects were sampled from each population

	Hearing	Deaf	Chinese
Young	1.69	0.98	1.34
Old	-2.97	-1.55	0.50

BPS - 5th Ed. Chapter 15 64

Case Study
Memory Loss

- Young Americans (hearing and deaf) have significantly higher mean scores.
- Science News (July 2, 1994, p. 13): "Surprisingly, ...memory scores for older and younger Chinese did not statistically differ."

BPS - 5th Ed. Chapter 15 65

Case Study
Memory Loss

- Since the sample sizes are very small, there is an increased chance that the test will result in no statistically significant difference being detected even if indeed there is a difference between young and old subjects' mean memory scores.
- The "surprising" result *could* just be because the sample size was too small to statistically detect a difference. A larger sample *may* yield different results.

BPS - 5th Ed. Chapter 15 66

Cautions About Significance Tests

Beware of Multiple Analyses

- Suppose that 20 null hypotheses are true.
- Each test has a 5% chance of being significant at the 5% level. That's what α = 0.05 means: results this extreme occur only 5% of the time just by chance when the null hypothesis is true.
- Thus, we expect about 1 in 20 tests (which is 5%) to give a significant result just by chance.
- Running one test and reaching the α = 0.05 level is reasonably good evidence against H_0; running 20 tests and reaching that level only once is not.

Similarly, the probability that all of twenty 95% confidence intervals will capture their true mean is much less than 95%.

BPS - 5th Ed. Chapter 15 67

Decision Errors: Type I

- If we reject H_0 when in fact H_0 is true, this is a **Type I error**.
- If we *decide* there is significant evidence in the data to reject the null hypothesis:
  - This is an incorrect decision if H_0 is in fact true.
  - The probability of this incorrect decision is equal to α.
- If the *null is true,* α = 0.05, and we rejected:
  - The extremity of the test statistic is due to chance.
  - About 5% of all samples from this population will lead us to wrongly reject chance and conclude significance.

BPS - 5th Ed. Chapter 15 68

Decision Errors: Type II

- If we fail to reject H_0 when in fact H_a is true, this is a **Type II error**.
- If we *decide* not to reject chance and thus allow for the plausibility of the null hypothesis
  - This is an incorrect decision if H_a is true.
  - The probability of NOT making this incorrect decision is called the power of the test.

BPS - 5th Ed. Chapter 15 69

Decision Errors: Type II

- Having a small significance level (low probability of Type I error) and a high power (low probability of Type II error) would be the best.
- Controlling both is often hard.
- Larger sample sizes help.
- If we can't control both, we opt for a small significance level, since Type I error involves making a decision (rejecting the null) while Type II error doesn't (not rejecting the null).

BPS - 5th Ed. Chapter 15 70

Decision Errors: Type I & Type II

		Truth about the population	
		H_0 true	H_a true
Decision based on sample	Reject H_0	Type I error	Correct decision
	Accept H_0	Correct decision	Type II error

BPS - 5th Ed. Chapter 15 71

Planning Studies

The Power of a Test

- The probability that a fixed level α significance test will reject H_0 when a particular alternative value of the parameter is true is called the **power** of the test against that specific alternative value.
- While α gives the probability of wrongly rejecting H_0 when in fact H_0 is true, power gives *the probability of correctly rejecting H_0 when in fact H_0 should be rejected* (because the value of the parameter is some specific value satisfying the alternative hypothesis)
- When μ is close to μ_0, the test will find it hard to distinguish between the two (low power); however, when μ is far from μ_0, the test will find it easier to find a difference (high power).

BPS - 5th Ed. Chapter 15 72

Case Study
Sweetening Colas (Ch. 14)

- The cola maker determines that a sweetness loss is too large to be acceptable if the mean response for all tasters is μ = 1.1 (or larger)
- Will a 5% significance test of the hypotheses
  H_0: $\mu = 0$ H_a: $\mu > 0$
  based on a sample of 10 tasters usually detect a change this great (rejecting H_0)?

BPS - 5th Ed. Chapter 15 73

Case Study
Sweetening Colas

1. Write the rule for rejecting H_0 in terms of $\bar{x}$.

   We know that σ = 1, so the *z* test rejects H_0 at the α = 0.05 level when

   $$z = \frac{\bar{x} - 0}{1/\sqrt{10}} \geq 1.645$$

   This is the same as:

   Reject H_0 when $\bar{x} \geq 0 + 1.645\frac{1}{\sqrt{10}} = 0.520$

   This step just restates the rule for the test. It pays no attention to the specific alternative we have in mind.

BPS - 5th Ed. Chapter 15 74

Case Study
Sweetening Colas

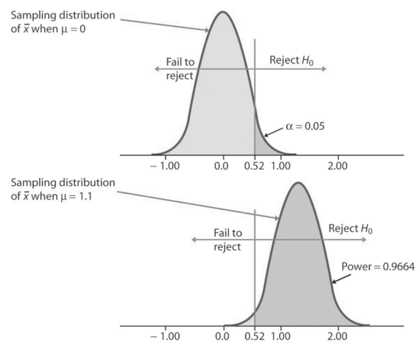
2. The power is the probability of rejecting H_0 under the condition that the alternative μ = 1.1 is true.

   To calculate this probability, standardize $\bar{x}$ using μ = 1.1 :

   $$P(\bar{x} \geq 0.520 \text{ when } \mu = 1.1) = P\left(\frac{\bar{x} - 1.1}{1/\sqrt{10}} \geq \frac{0.520 - 1.1}{1/\sqrt{10}}\right)$$
   $$= P(Z \geq -1.83) = 1 - 0.0336 = 0.9664$$

   96.64% of tests will declare that the cola loses sweetness when the true mean sweetness loss is 1.1 (power = 0.9664).

BPS - 5th Ed. Chapter 15 75



BPS - 5th Ed. Chapter 15 76