Experimental Design
Note 2
Basic Statistical Methods

Keunbaik Lee

Sungkyunkwan University

Statistics: two major chapters

- Descriptive Statistics
  - Gives numerical and graphic procedures to summarize a collection of data in a clear and understandable way
- Inferential Statistics
  - Provides procedures to draw inferences about a population from a sample

Some Basic Statistical Concepts

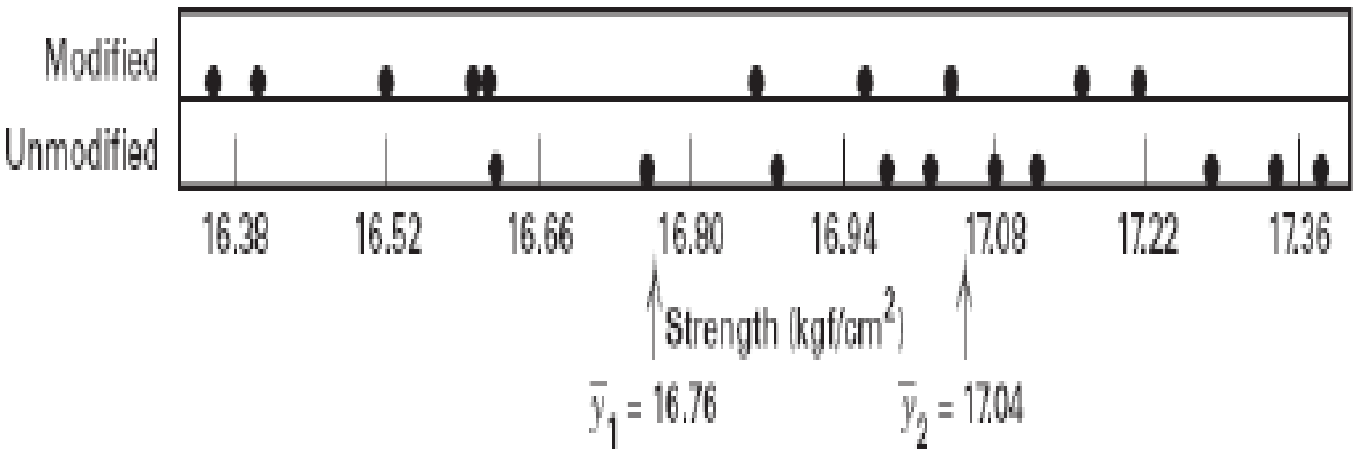
- Describing sample data
  - Random samples
  - Sample mean, variance, standard deviation
  - Populations versus samples
  - Population mean, variance, standard deviation
  - Estimating parameters
- Simple comparative experiments
  - The hypothesis testing framework
  - The two-sample t-test
  - Checking assumptions, validity

Portland Cement Formulation (Page 26)

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

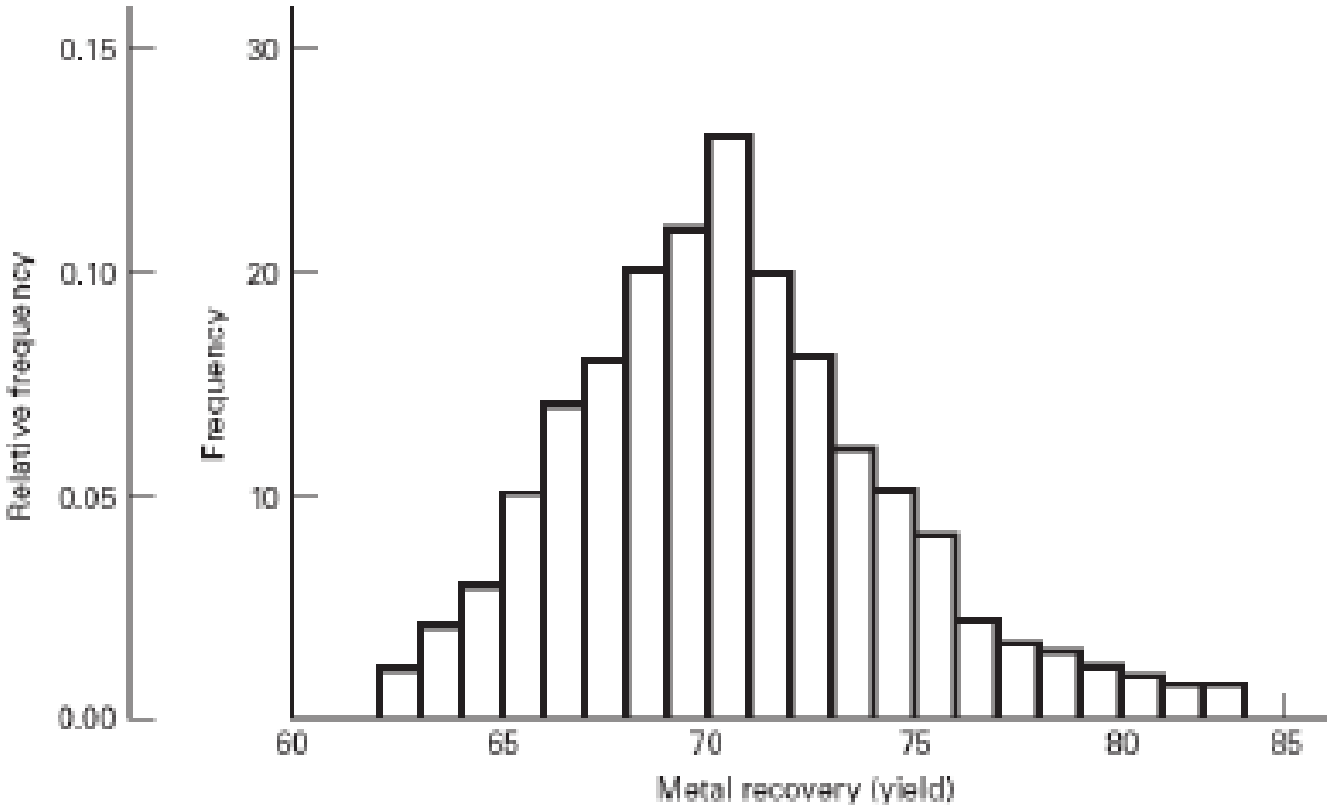
j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Graphical View of the Data: dot diagram



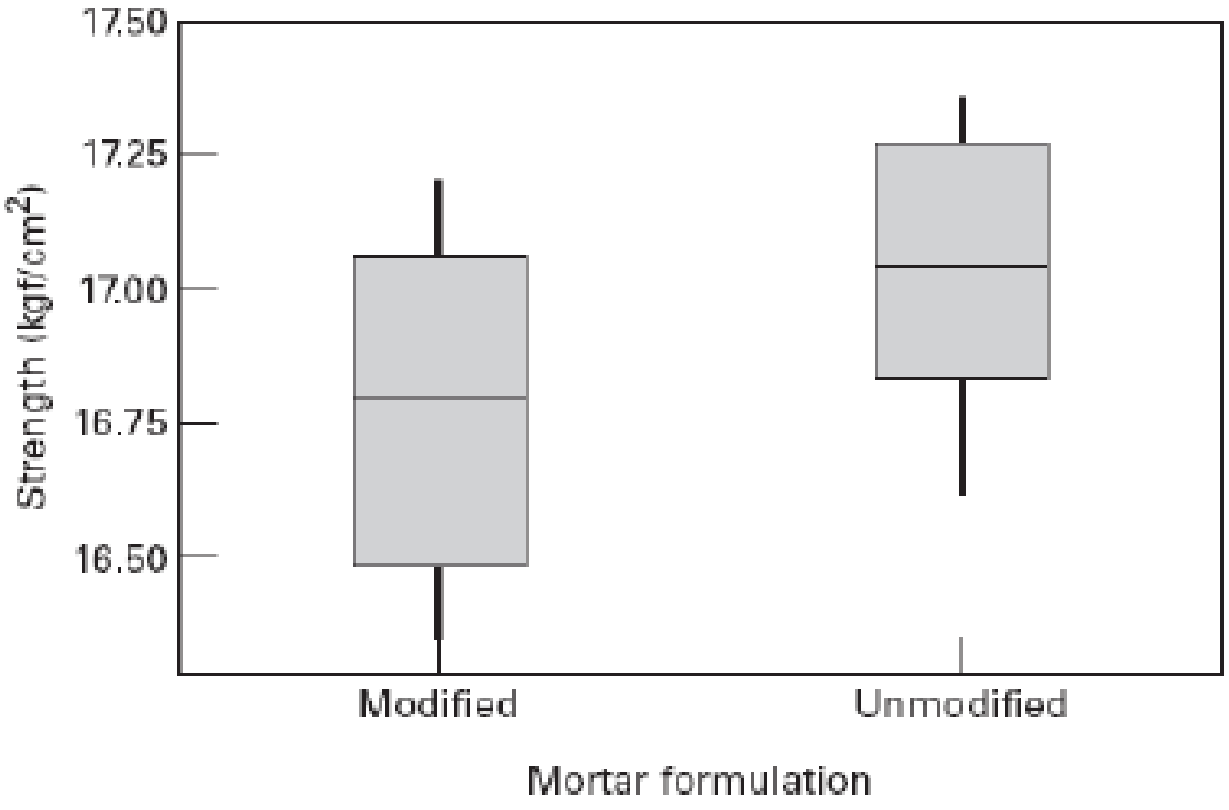
■ FIGURE 2.1 Dot diagram for the tension bond strength data in Table 2.1

If you have a large sample, a histogram may be useful



■ **FIGURE 2.2** Histogram for 200 observations on metal recovery (yield) from a smelting process

Box Plots



■ **FIGURE 2.3** **Box plots for the Portland cement tension bond strength experiment**

Numerical approach - Descriptive Measures

- Central tendency measures compute to give a "center" around which the measurements in the data are distributed
- Variation or Variability measures describe data spread or how far away the measurements are from the center
- Skewness describe how much the data is skewed to one side
- Kurtosis measure the peak of the data

Random Variable and Probability Distribution

- Discrete random variable Y
  - Finite possible values $\{y_1, y_2, \cdots, y_k\}$
  - Probability mass function $\{p(y_1), p(y_2), \cdots, p(y_k)\}$ satisfying

$$p(y_i) \geq 0$$

and

$$\sum_{i=1}^{k} p(y_i) = 1$$

- Continuous random variable Y
  - Possible values from an interval
  - Probability density function $f(y)$ satisfying

$$f(y) \geq 0$$

and

$$\int f(y)dy = 1$$

Mean, variance, formulas

- Mean and variance
  - Mean $\mu = E(Y)$: center, location etc.
  - Variance $\sigma^2 = var(Y)$: spread, dispersion etc.
  - Discrete Y:

$$\mu = \sum_{i=1}^{k} y_i p(y_i); \quad \sigma^2 = \sum_{i=1}^{k} (y_i - \mu)^2 p(y_i)$$

  - Continuous Y:

$$\mu = \int y f(y) dy; \quad \sigma^2 = \int (y - \mu)^2 f(y) dy$$

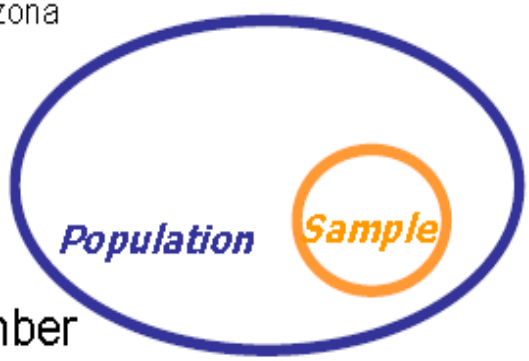
- Formulas for calculating mean and variance for two variables
  - If Y_1 and Y_2 are independent, then

$$E(Y_1 Y_2) = E(Y_1)E(Y_2); \quad var(aY_1 \pm bY_2) = a^2 var(Y_1) + b^2 var(Y_2)$$

Population vs Sample

- **Population:** The entire group of individuals in which we are interested but can't assess directly
  - Example: all adult males in Arizona

- **Sample:** A part of the population we actually examine and for which we do have data
  - Example: Take a sample of 100 adult males in Arizona



How well the sample represents the population depends on the *sampling design.*

- A **parameter** is a number describing a characteristic of the population.
  - Example: average height of the adult males in Arizona
  - Usually parameter is unknown and need to be estimated by using ----

- A **statistic** is a number describing a characteristic of a sample.
  - –Example: average height of these 100 adult males
  - –Different sample results in different statistic

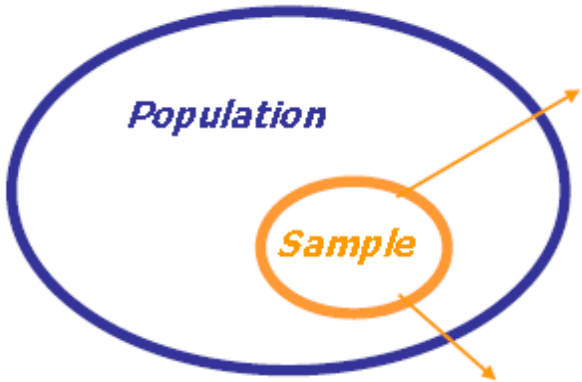
Parameter - Statistic - Estimate

- Parameter, which is unknown, is our interest
- A statistic, distinct from an unknown parameter, can be computed from a sample
- Very often, a statistic used to estimate a parameter is also called an estimate
  - For instance, the sample mean is a statistic and is an estimate for the population mean, which is a parameter

Mean and standard deviation from a population and from a sample

- **Population:**
  - Population mean μ
  - Population standard deviation σ

- **Sample:**
  - Sample mean



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  - Sample standard deviation: variation around the mean

$$s = \sqrt{\frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x})^2}$$

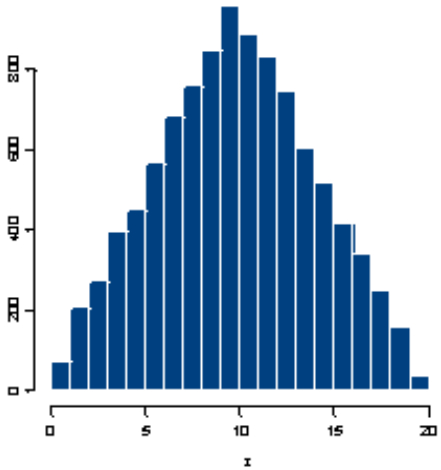
Why sampling distribution?

- We have discussed (point) estimates:
  - as an estimate of population mean, μ
  - as an estimate of population standard deviation, σ
- These (point) estimates are almost never exactly equal to the true values they are estimating
- In order for the point estimate to be useful, it is necessary to describe just how far off from the true value it is likely to be

Sampling Distributions I

- By drawing random samples of size N from a population with a specific mean and variance, we can learn
  - how much error we can expect on average and
  - how much variation there will be on average in the errors observed
- Sampling distribution: the distribution of a sample statistic (e.g., a mean) when sampled under known sampling conditions from a known population

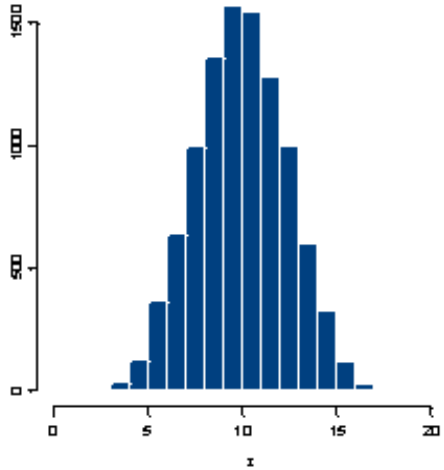
Sampling Distributions II



n = 2

mean of sample means = 10

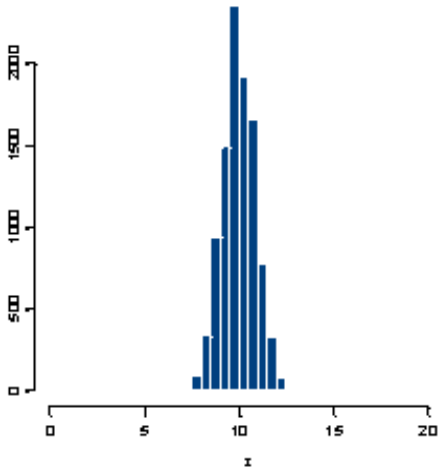
SD of sample means = 4.16



n = 5

mean of sample means = 10

SD of sample means = 2.41

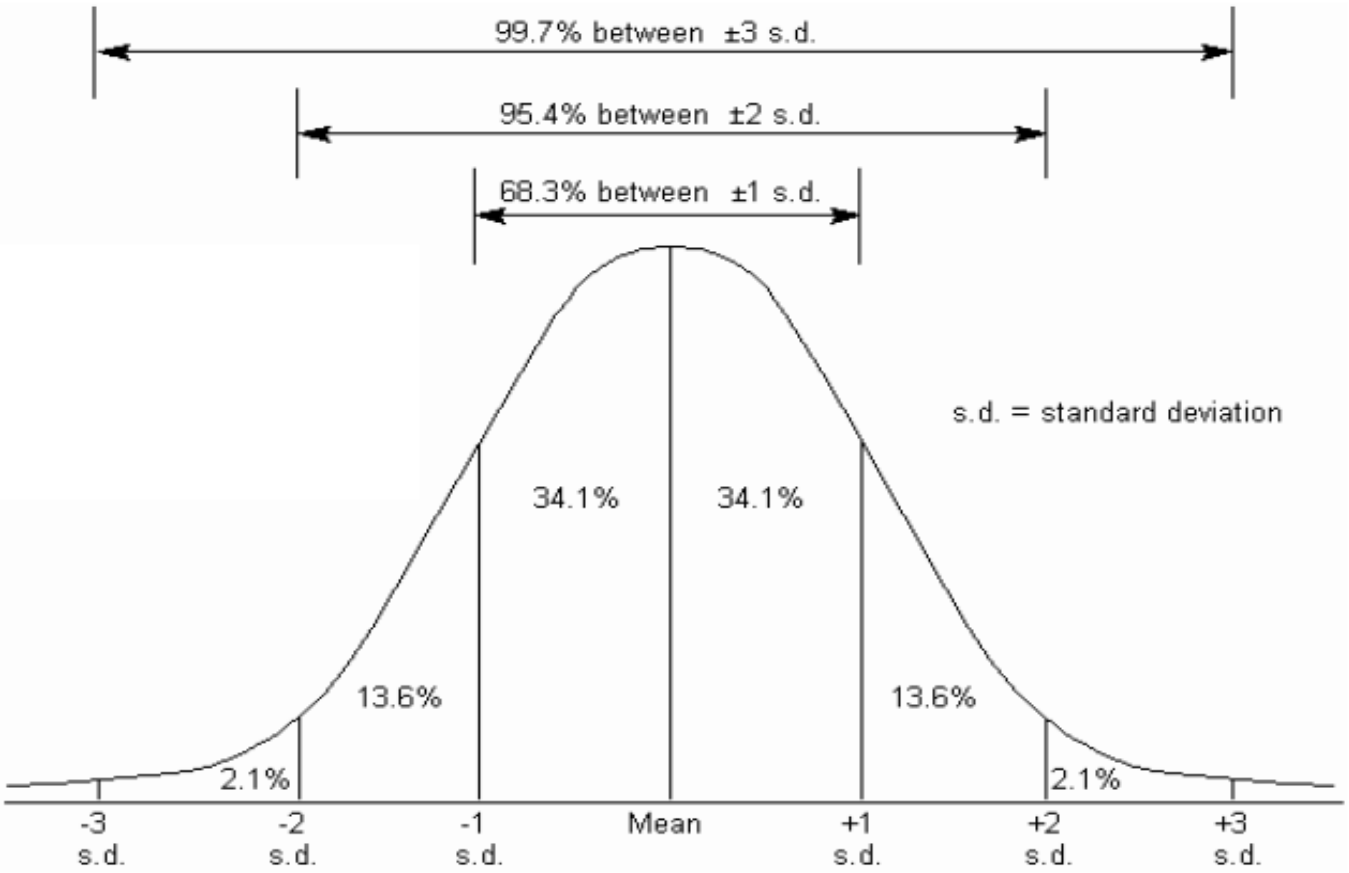


n = 15

mean of sample means = 10

SD of sample means = 0.87

Normal Distributions I



Normal Distributions II

Random sample: $Y_1, Y_2, \cdots, Y_n \sim^{iid} N(\mu, \sigma^2)$
Sample mean

$$\bar{Y} = (Y_1 + Y_2 + \cdots + Y_n)/n$$
$$E(\bar{Y}) = \mu, \quad var(\bar{Y}) = \sigma^2/n$$

$\bar{Y}$ follows $N(\mu, \sigma^2/n)$.

Central Limit Theorem I

$Y_1, Y_2, \cdots, Y_n$ are n independent and identically distributed random variables with $E(Y_i) = \mu$ and $var(Y_i) = \sigma^2$. Then

$$Z_n = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

approximately follows the standard normal distribution $N(0, 1)$.

Remark:

- Do not need to assume the original population distribution is normal.
- When the population distribution is normal, then Z_n exactly follows $N(0, 1)$.

Sampling distribution: sample variance I

$Y_1, Y_2, \cdots, Y_n \sim^{iid} N(\mu, \sigma^2)$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2,$$

$$E(S^2) = \sigma^2,$$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Chi-squared distribution

If $Z_1, Z_2, \cdots, Z_k$ are i.i.d. as $N(0,1)$, then

$$W = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

follows a Chi-squared distribution with degree of freedom (df) k, denoted by χ_k^2

Density functions of χ_k^2

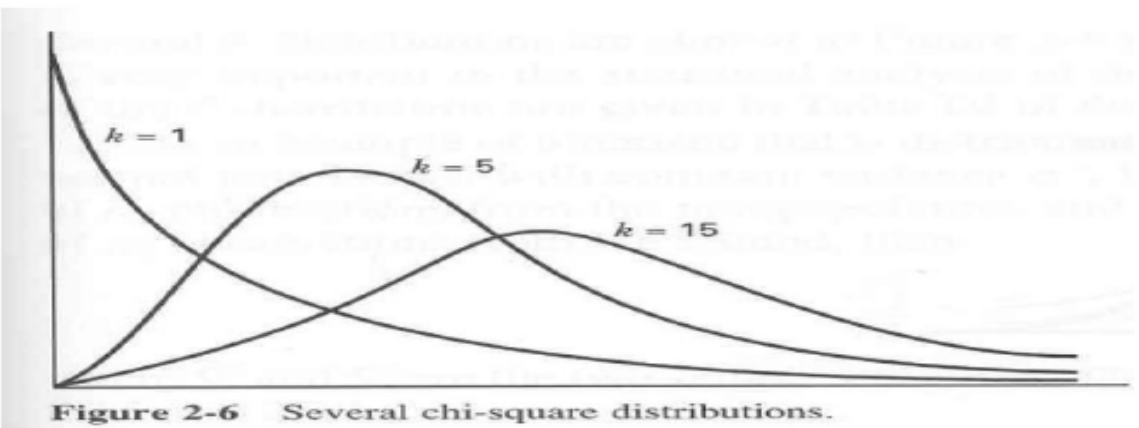


Figure 2-6 Several chi-square distributions.

t-distribution I

If $Z \sim N(0,1)$, $W \sim \chi^2_k$, and Z and W are independent, then

$$T_k = \frac{Z}{\sqrt{W/k}}$$

follows a t-distribution with d.f. k, denoted by $t_{(k)}$. For example, in t-test:

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} = \frac{\sqrt{n}(\bar{Y} - \mu_0)/\sigma}{\sqrt{S^2/\sigma^2}} = \frac{Z}{\sqrt{W/(n-1)}} \sim t_{(n-1)}.$$

Remark:
As n goes to ∞, $t_{(n-1)}$ converges to $N(0,1)$.

t-distribution II

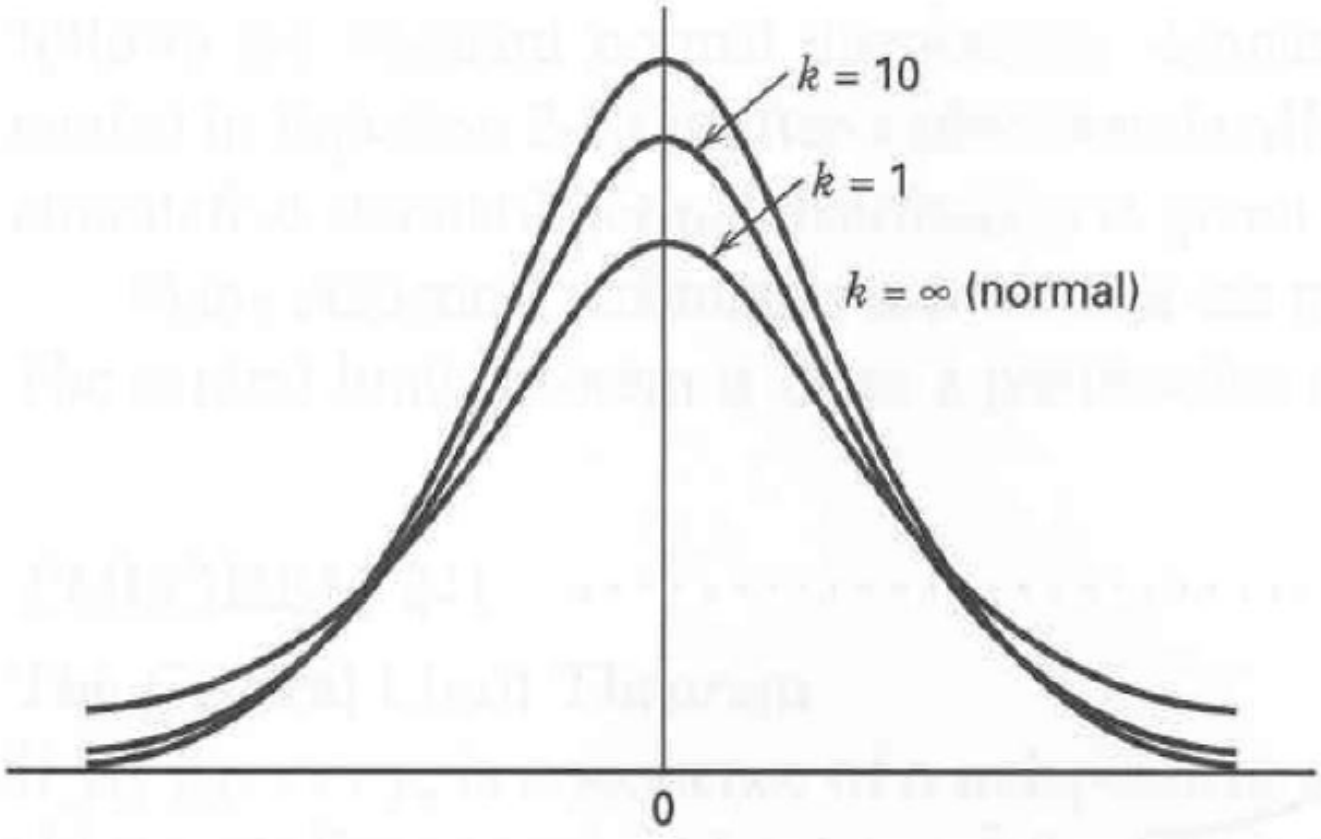


Figure 2-7 Several t distributions.

F-distribution I

- F-distributions: F_{k_1,k_2}
  Suppose random variables $W_1 \sim \chi^2_{k_1}$, $W_2 \sim \chi^2_{k_2}$, and W_1 and W_2 are independent, then

$$F = \frac{W_1/k_1}{W_2/k_2}$$

  follows F_{k_1,k_2} with numerator of d.f. k_1 and denominator d.f. k_2.

F-distribution II

- Example: $H_0 : \sigma_1^2 = \sigma_2^2$, the test statistic is

$$F = \frac{S_1^2}{S_2^2} = \frac{S_1^2/\sigma^2}{S_2^2/\sigma^2} = \frac{W_1/(n_1-1)}{W_2/(n_2-1)} \sim F_{n_1-1,n_2-1}.$$

  Refer to Section 2.3 for details.

F-distribution III

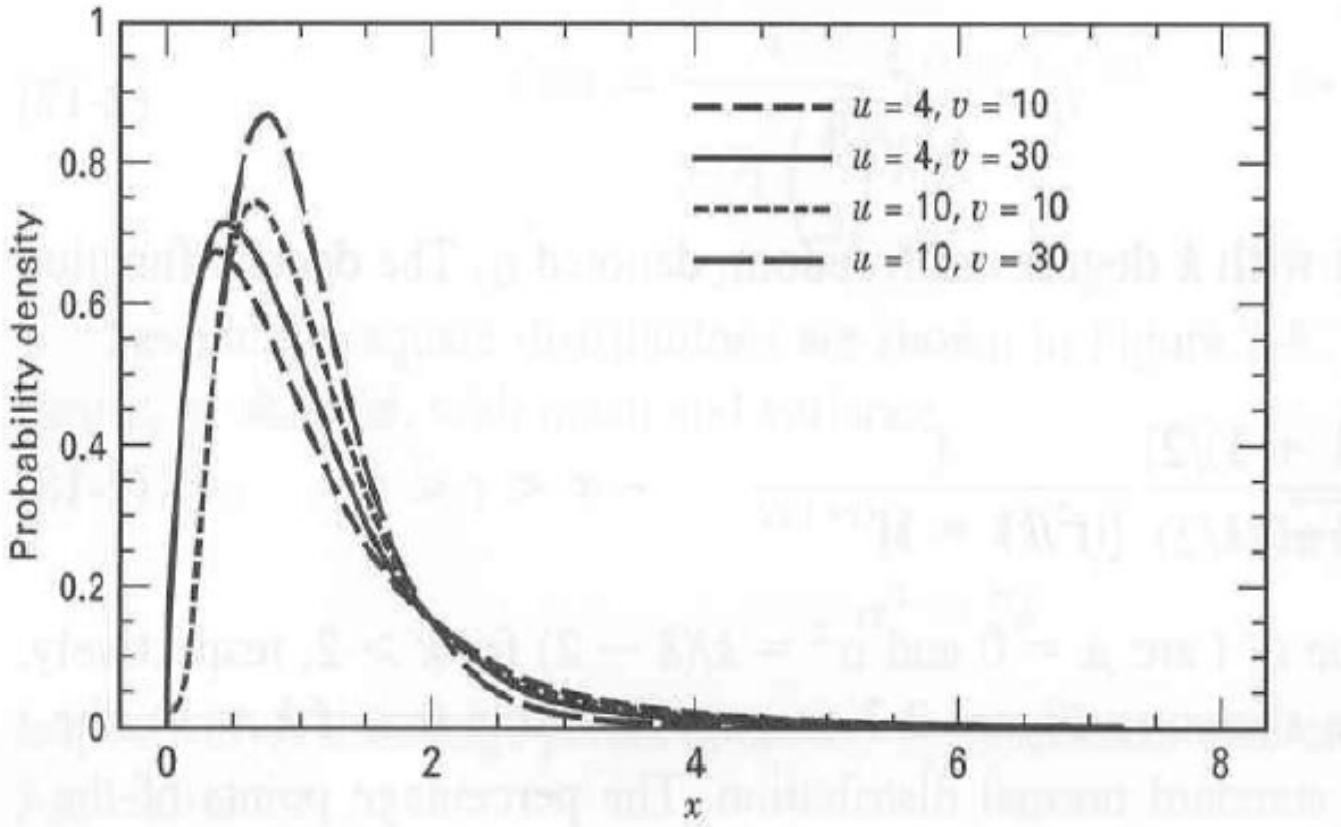


Figure 2-8 Several F distributions.

Statistical Inference - Hypothesis Testing I

- Statistical hypothesis testing is a useful framework for many experimental situations
- Origins of the methodology date from the early 1900s
- We will use a procedure known as the two-sample t-test

Inference about the Differences in Means I

- Hypothesis Testing

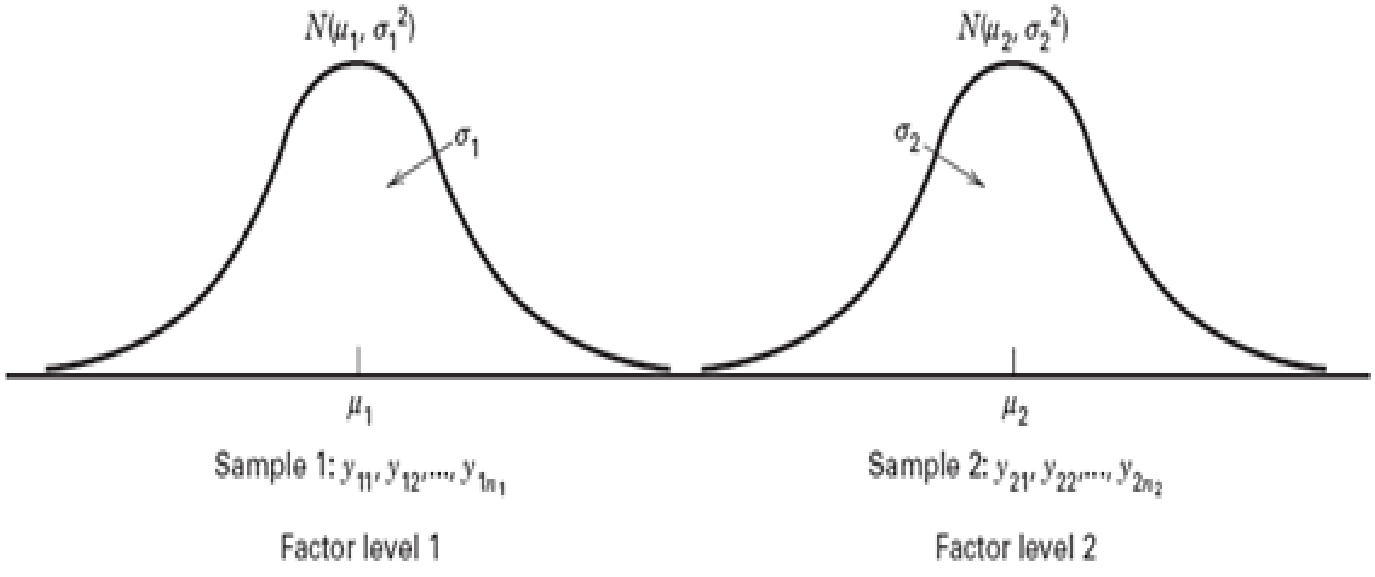


FIGURE 2.9 The sampling situation for the two-sample t-test

Inference about the Differences in Means II

- Two-sample t-test:
  Sampling from a normal distribution
  Statistical hypotheses: $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$
  Test statistic:

$$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

  In Example,

$$\bar{y}_1 = 16.76,\ S_1^2 = 0.100,\ S_1 = 0.316,\ n_1 = 10$$
$$\bar{y}_2 = 17.04,\ S_2^2 = 0.061,\ S_1 = 0.248,\ n_2 = 10$$

Inference about the Differences in Means III

- Two sample pooled t-test: if $\sigma^2 = \sigma_1^2 = \sigma_2^2$ assumed unknown

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

  - Values of t_0 that are near zero are consistent with the null hypothesis
  - Values of t_0 that are very different from zero are consistent with the alternative hypothesis
  - t_0 is a "distance" measure-how far apart the averages are expressed in standard deviation units
  - Notice the interpretation of t_0 as a signal-to-noise ratio

Inference about the Differences in Means IV

In Example,

$$S_p^2 = \frac{9(0.100) + 9(0.061)}{10 + 10 - 2} = 0.081 : \quad \text{pooled estimate}$$

$$S_p = 0.284$$

$$t = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}} = -2.20$$

The two sample means are a little over two standard deviations apart. Is this a "large" difference?

- William Sealy Gosset (1976 - 1937)

Inference about the Differences in Means V

Gosset's interest in barley cultivation led him to speculate that design of experiments should aim, not only at improving the average yield, but also at breeding varieties whose yield was insensitive (robust) to variation in soil and climate.

Developed the *t*-test (1908) using his pen name "student"

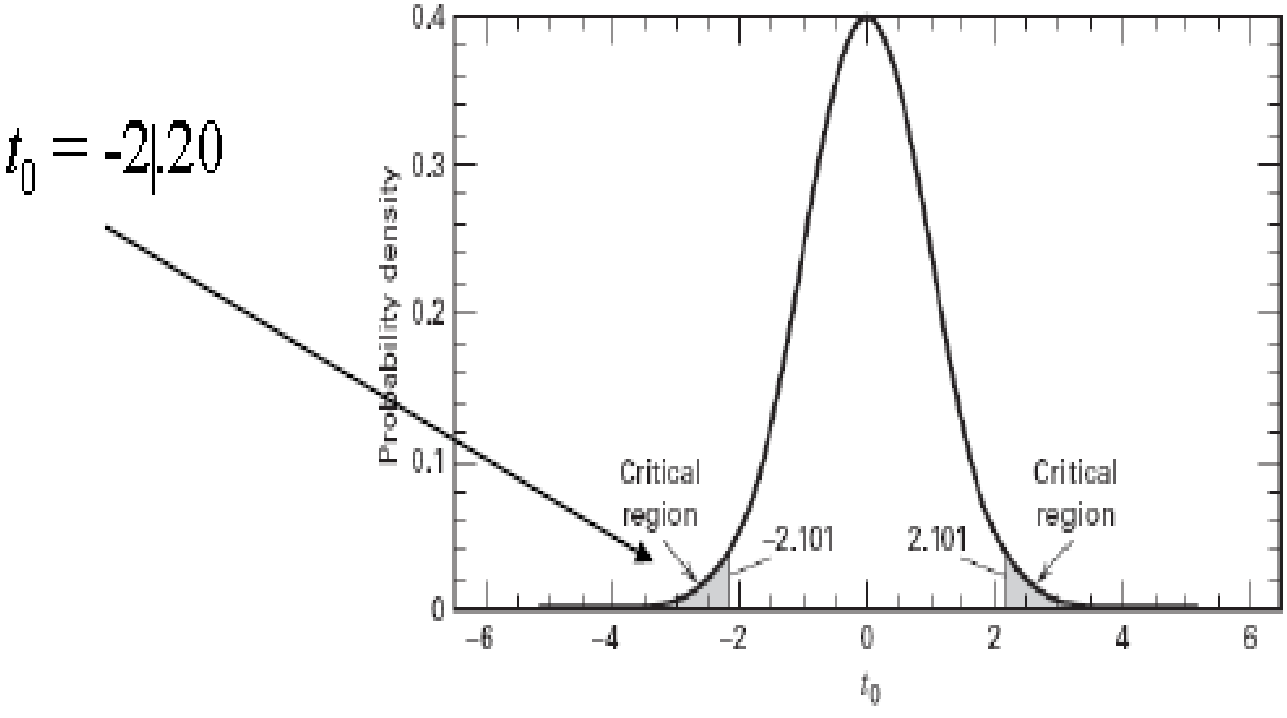
Gosset was a friend of both Karl Pearson and R.A. Fisher, an achievement, for each had a monumental ego and a loathing for the other.

'Student' in 1908

Inference about the Differences in Means VI

- Two-sample (pooled) t-Test
  - We need an objective basis for deciding how large the test statistic t_0 really is.
  - In 1908, W. S. Gosset derived the reference distribution for t_0, called the t distribution
  - Tables of the t distribution - see textbook appendix
  - The P-value is the area (probability) in the tails of the t-distribution beyond -2.20 + the probability beyond +2.20 (its a two-sided test)
  - The P-value is a measure of how unusual the value of the test statistic is given that the null hypothesis is true
  - The P-value is the risk of wrongly rejecting the null hypothesis of equal means (it measures rareness of the event)
  - The P-value in our problem is $P = 0.042$

Inference about the Differences in Means VII



■ **FIGURE 2.10** The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$

Inference about the Differences in Means VIII

- Given significance level α, there are two approaches:
  - Compare observed test statistic with critical value
  - Compute the P-values of observed test statistic
    Reject H_0, if the P-value $\leq \alpha$

Two types of error I

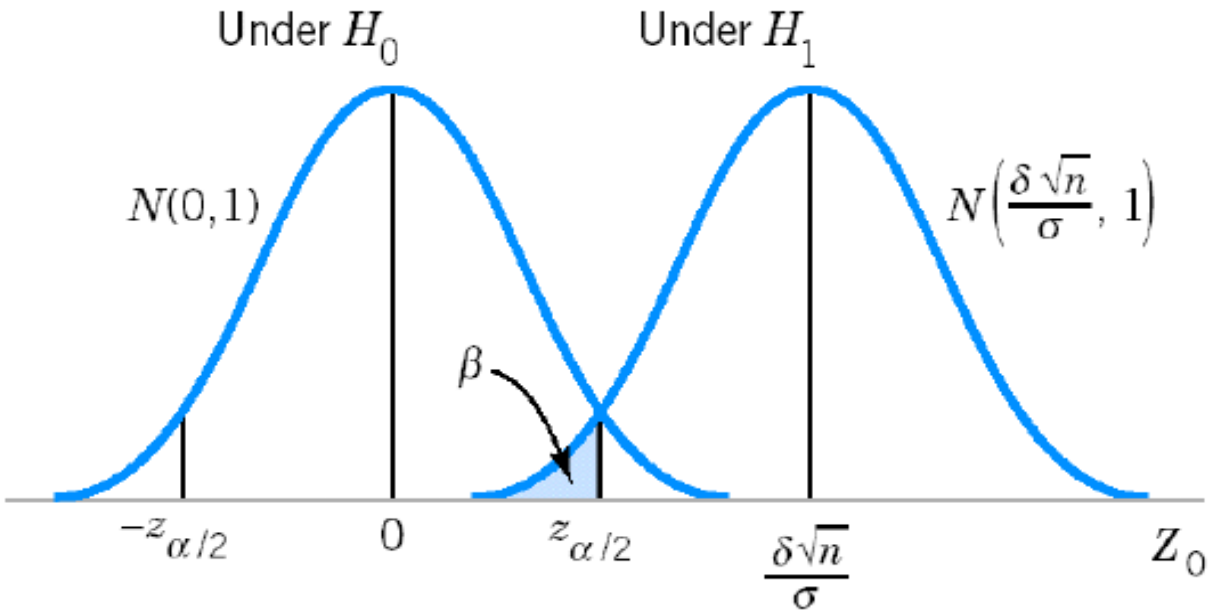
Truth / Data	H_0 Correct (no disease)	H_1 Correct (disease)
Decide H_0 "fail to reject H_0" (test shows no disease)	1- α True Negative	β False Negative
Decide H_1 "reject H_0" (test shows disease)	α False Positive	1- β True Positive

- Type I error - False positive rate
- $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$
  - Probability reject the true null hypothesis
- α is significance level

- Type II error--False negative rate
- $\beta = P(\text{do not reject } H_0 \mid H_1 \text{ true})$
  - Probability not reject a false null hypothesis
- **Power** = $1-\beta = P(\text{reject } H_0 \mid H_1 \text{ true})$

Two types of error II

- Type I, II error
  - $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, and assume variance σ^2 is known.
  - Let $\delta = \mu - \mu_0$.
  - Let $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, then under $H_0 : Z_0 \sim N(0,1)$ and under H_1



Two types of error III

- Type II error depend on: α, sample size, population variance, difference between actual and hypothesized means
- To decrease β (type II error) for a fixed α (type I error), we need increase sample size n.

- Why we have such distributions in the previous slide? Consider the two-sided hypothesis $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, say, where $\delta > 0$. The test statistic Z_0 is

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \frac{\delta\sqrt{n}}{\sigma}.$$

Two types of error IV

Therefore, the distribution of Z_0 when H_1 is true is

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right)$$

- Choice of sample size (one sample)

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

$$n \simeq \frac{(z_{\alpha/2} + z_{\beta})^2\sigma^2}{\delta^2}$$

Two types of error V

- Choice of sample size (two sample case, known variances)
  $H_0 : \mu_1 = \mu_2 + \Delta_0$ vs $H_1 : \mu_1 \neq \mu_2 + \Delta_0$
  - Assume two samples have the same size n, then

$$n \simeq \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

    where Δ is the true difference in means.
  - An alternative way for calculating sample size is operating characteristic curve (read p44-46).
- For one-sided test of two-sample case,

$$n = \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}.$$

Example I

- A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested: formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce drying time.
- From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient.
- Suppose that if the true difference in drying times is as much as 10 minutes. He wants to detect this with probability at least 0.90.

Example II

- Under the null hypothesis $\Delta_0 = 0$ and one-sided alternative hypothesis is with $\Delta = 10$. Since the power is 0.9, $\beta = 0.10$ ($Z_\beta = Z_{0.10} = 1.28$). Given $\alpha = 0.05$ (so $Z_\alpha = 1.645$),

$$n = \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta)^2} = \frac{(1.645 + 1.28)^2(8^2 + 8^2)}{(10 - 0)^2} = 11$$

Summary of Tests I

■ TABLE 2.3

Tests on Means with Variance Known

	Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection	*P*-Value
One sample	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$\lvert Z_0 \rvert > Z_{\alpha/2}$	$P = 2[1 - \Phi(\lvert Z_0 \rvert)]$
	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$Z_0 = \dfrac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$	$Z_0 < -Z_\alpha$	$P = \Phi(Z_0)$
	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$Z_0 > Z_\alpha$	$P = 1 - \Phi(Z_0)$
two samples	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$		$\lvert Z_0 \rvert > Z_{\alpha/2}$	$P = 2[1 - \Phi(\lvert Z_0 \rvert)]$
	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$Z_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$Z_0 < -Z_\alpha$	$P = \Phi(Z_0)$
	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$		$Z_0 > Z_\alpha$	$P = 1 - \Phi(Z_0)$

Summary of Tests II

TABLE 2.4
Tests on Means of Normal Distributions, Variance Unknown

	Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection	*P*-Value
One sample	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$\lvert t_0 \rvert > t_{\alpha/2,n-1}$	sum of the probability above t_0 and below $-t_0$
	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$t_0 = \dfrac{\bar{y} - \mu_0}{S/\sqrt{n}}$	$t_0 < -t_{\alpha,n-1}$	probability below t_0
	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$t_0 > t_{\alpha,n-1}$	probability above t_0
two samples		if $\sigma_1^2 = \sigma_2^2$		
	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	$\lvert t_0 \rvert > t_{\alpha/2,v}$	sum of the probability above t_0 and below $-t_0$
		$v = n_1 + n_2 - 2$		
		if $\sigma_1^2 \neq \sigma_2^2$		
	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	$t_0 < -t_{\alpha,v}$	probability below t_0
	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$v = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}$	$t_0 > t_{\alpha,v}$	probability above t_0

Summary of Tests III

■ TABLE 2.7

Tests on Variances of Normal Distributions

	Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection
One sample	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$		$\chi_0^2 > \chi^2_{\alpha/2,n-1}$ or $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$
	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	$\chi_0^2 < \chi^2_{1-\alpha,n-1}$
	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$		$\chi_0^2 > \chi^2_{\alpha,n-1}$
two samples	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$F_0 = \dfrac{S_1^2}{S_2^2}$	$F_0 > F_{\alpha/2,n_1-1,n_2-1}$ or $F_0 < F_{1-\alpha/2,n_1-1,n_2-1}$
	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$	$F_0 = \dfrac{S_2^2}{S_1^2}$	$F_0 > F_{\alpha,n_2-1,n_1-1}$
	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$F_0 = \dfrac{S_1^2}{S_2^2}$	$F_0 > F_{\alpha,n_1-1,n_2-1}$

Confidence Intervals I

- Hypothesis testing gives an objective statement concerning the difference in means, but it does not specify "how different" they are.
- General form of a confidence interval

$$L \leq \theta \leq U \text{ where } P(L \leq \theta \leq U) = 1 - \alpha$$

- The $100(1-\alpha)\%$ confidence interval on the difference in two means:

$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2,n_1+n_2-2} S_p \sqrt{1/n_1 + 1/n_2} \leq \mu_1 - \mu_2 \leq$$
$$\bar{y}_1 - \bar{y}_2 + t_{\alpha/2,n_1+n_2-2} S_p \sqrt{1/n_1 + 1/n_2}$$

Checking normal assumptions I

- There are two ways of testing normality.
  - Graphical methods:
    visualize the distributions of random variables or differences between an empirical distribution and a theoretical distribution (e.g., the standard normal distribution).
  - Numerical methods:
    present summary statistics such as skewness and kurtosis, or conduct statistical tests of normality.

Checking normal assumptions II

	Graphical Methods	Numerical Methods
Descriptive	Stem-and-leaf plot, (skeletal) box plot, dot plot, histogram	Skewness Kurtosis
Theory-driven	P-P plot Q-Q plot	Shapiro-Wilk, Shapiro- Francia test Kolmogorov-Smirnov test (Lillefors test) Anderson-Darling/Cramer-von Mises tests Jarque-Bera test, Skewness-Kurtosis test

Graphical methods are intuitive and easy to interpret, while numerical methods provide objective ways of examining normality.

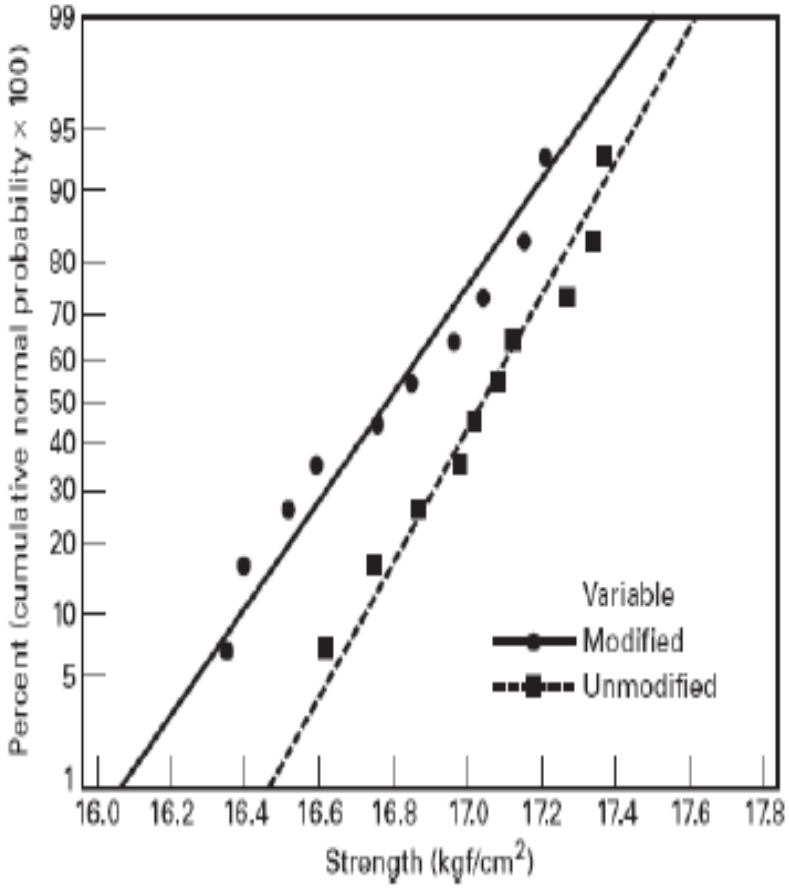
Checking normal assumptions III

- $Y_1, Y_2, \cdots, Y_n$ is a random sample from a population with mean μ and variance σ^2.
  - Order Statistic: $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$ where $Y_{(i)}$ is the ith smallest value.
  - If the population is normal, i.e., $N(\mu, \sigma^2)$, then

$$E(Y_{(i)}) \approx \mu + \sigma r_{\alpha_i} \text{ with } \alpha_i = \frac{i - 3/8}{n + 1/4}$$

    where r_{α_i} is the $100\alpha_i$th percentile of $N(0, 1)$ for $1 \leq i \leq n$.
  - Given a sample $y_1, y_2, \cdots, y_n$, the plot of $(r_{\alpha_i}, y_{(i)})$ is called the normal probability plot

Checking normal assumptions IV



■ **FIGURE 2.11** Normal probability plots of tension bond strength in the Portland cement experiment

Checking normal assumptions V

- the points falling around a straight line indicate normality of the population;
- Deviation from a straight line pattern indicates non-normality