Seminar in Statistics and methodology

Wednesday, 12 March 2008

Summary
Hypothesis testing
T-statistics

Eleonora Rossi
e.rossi@rug.nl

THE STANDARD DEVIATION s

The **variance** s^2 of a set of observations is the average of the squares of the deviations of the observations from their mean. In symbols, the variance of n observations $x_1, x_2, \ldots, x_n$ is

$$s^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n - 1}$$

or, in more compact notation,

$$s^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2$$

The **standard deviation** s is the square root of the variance s^2:

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

Definition, pg 49
Introduction to the Practice of Statistics, Fifth Edition

The normal distribution

- The density curve of a normal distribution is

  - Symmetric
  - Unimodal
  - Bell-shaped

  The flatness of the curve will depend on the SD

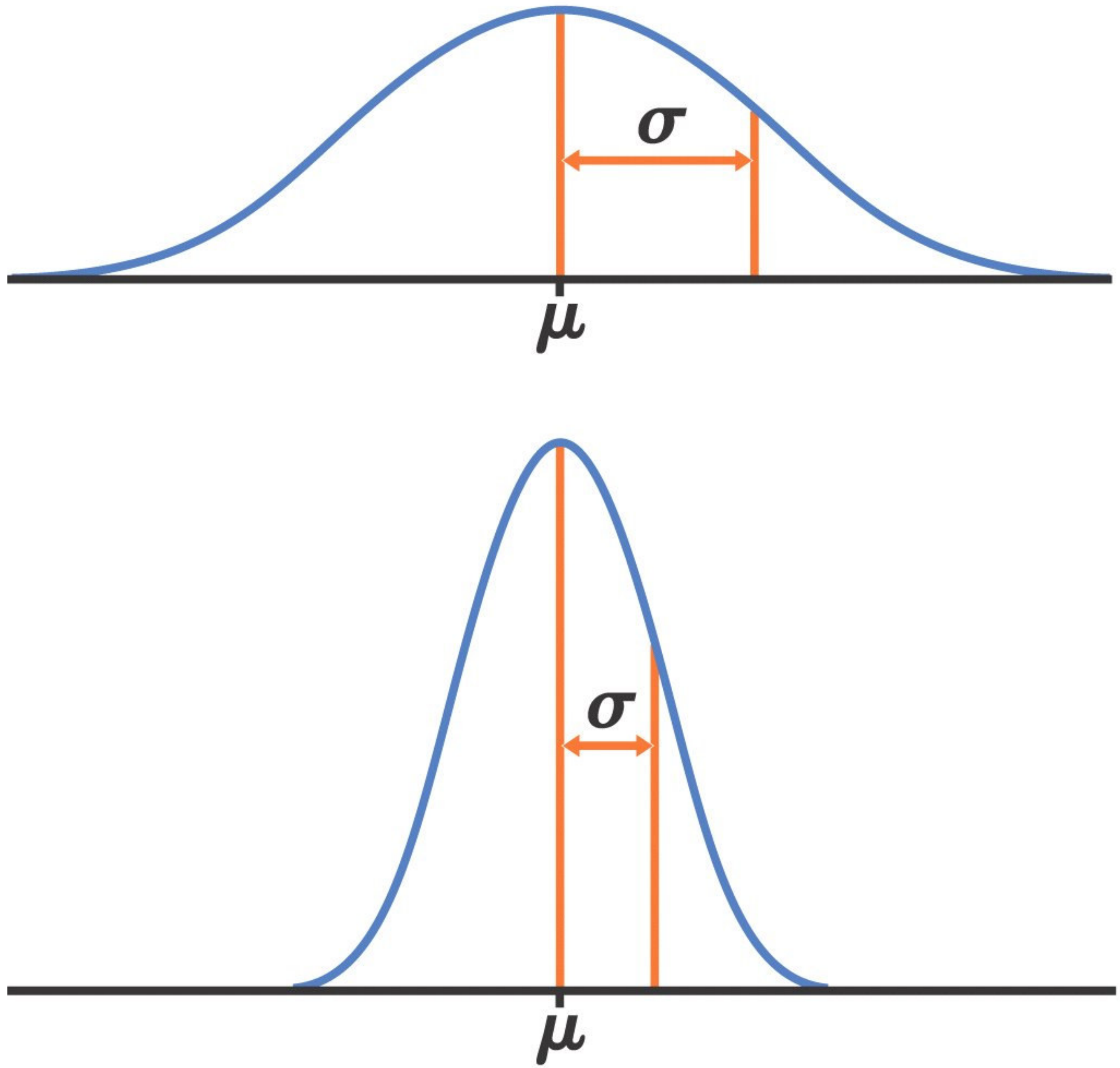


Figure 1-26
Introduction to the Practice of Statistics, Fifth Edition

THE 68–95–99.7 RULE

In the normal distribution with mean μ and standard deviation σ:

- Approximately **68%** of the observations fall within σ of the mean μ.
- Approximately **95%** of the observations fall within 2σ of μ.
- Approximately **99.7%** of the observations fall within 3σ of μ.

Definition, pg 71
Introduction to the Practice of Statistics, Fifth Edition

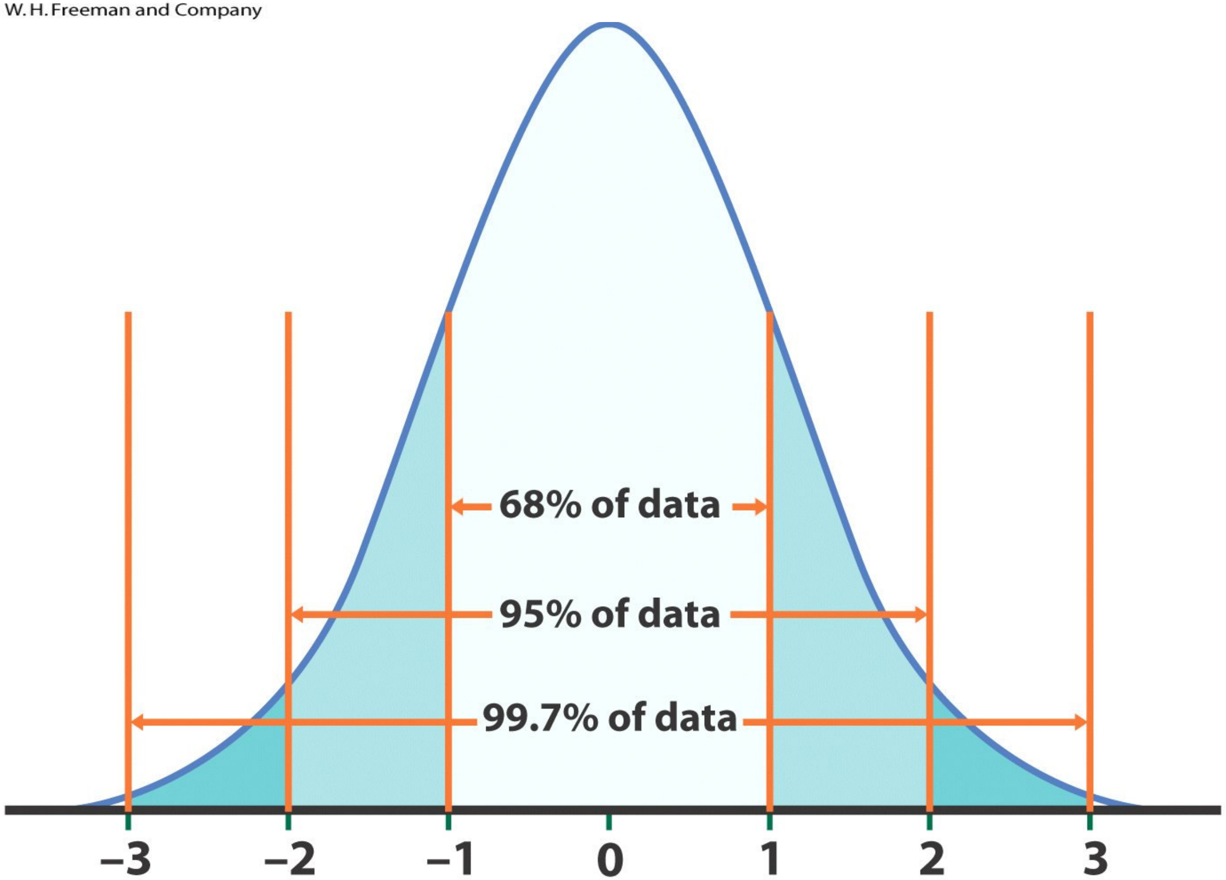


Figure 1-27
Introduction to the Practice of Statistics, Fifth Edition

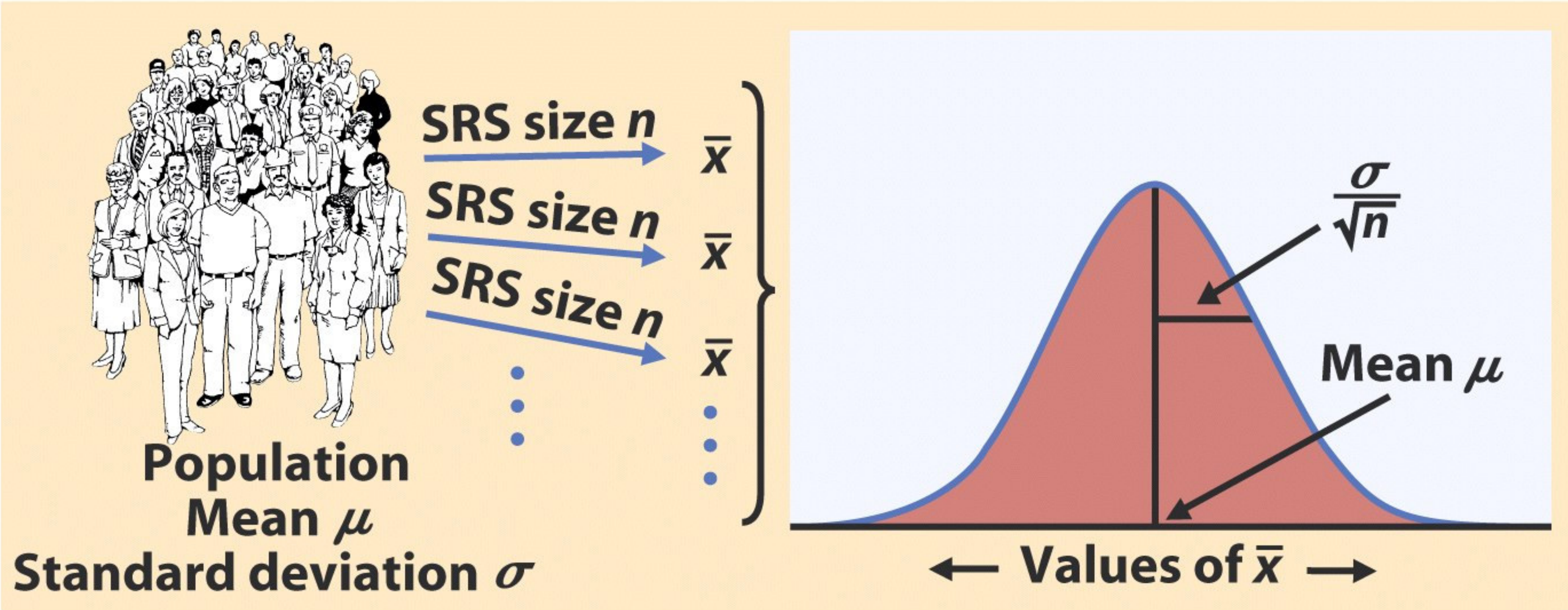


Figure 5-12

CENTRAL LIMIT THEOREM

Draw an SRS of size n from any population with mean μ and finite standard deviation σ. When n is large, the sampling distribution of the sample mean $\overline{x}$ is approximately normal:

$$\overline{x} \text{ is approximately } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Definition, pg 362b

Introduction to the Practice of Statistics, Fifth Edition

CONFIDENCE INTERVAL

A level C **confidence interval** for a parameter is an interval computed from sample data by a method that has probability C of producing an interval containing the true value of the parameter.

Definition, pg 387
Introduction to the Practice of Statistics, Fifth Edition

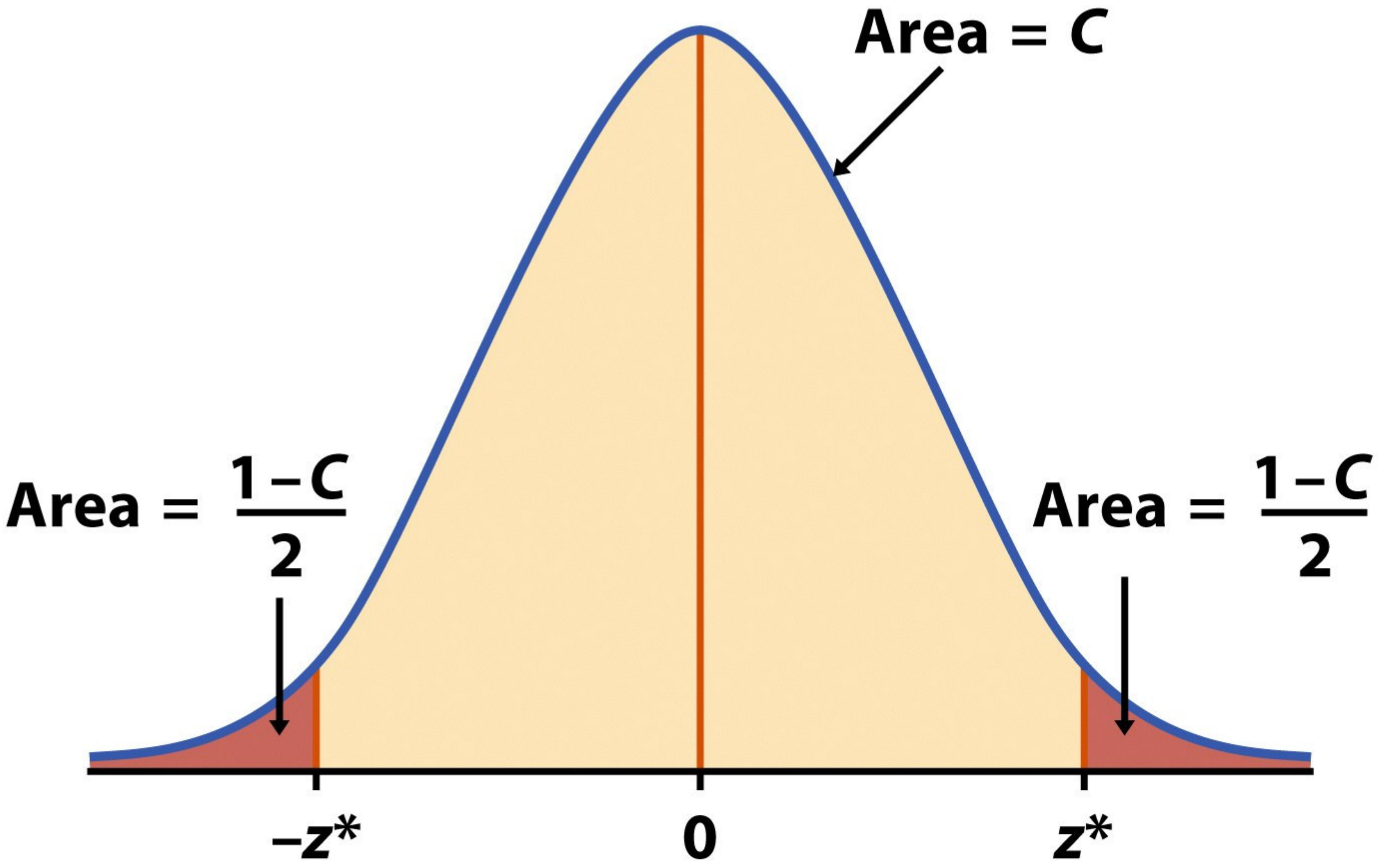


Figure 6-4
Introduction to the Practice of Statistics, Fifth Edition

CONFIDENCE INTERVAL FOR A POPULATION MEAN

Choose an SRS of size n from a population having unknown mean μ and known standard deviation σ. The **margin of error** for a level C confidence interval for μ is

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

Here z^* is the value on the standard normal curve with area C between the critical points $-z^*$ and z^*. The level C **confidence interval** for μ is

$$\bar{x} \pm m$$

This interval is exact when the population distribution is normal and is approximately correct for large n in other cases.

Definition, pg 388
Introduction to the Practice of Statistics, Fifth Edition

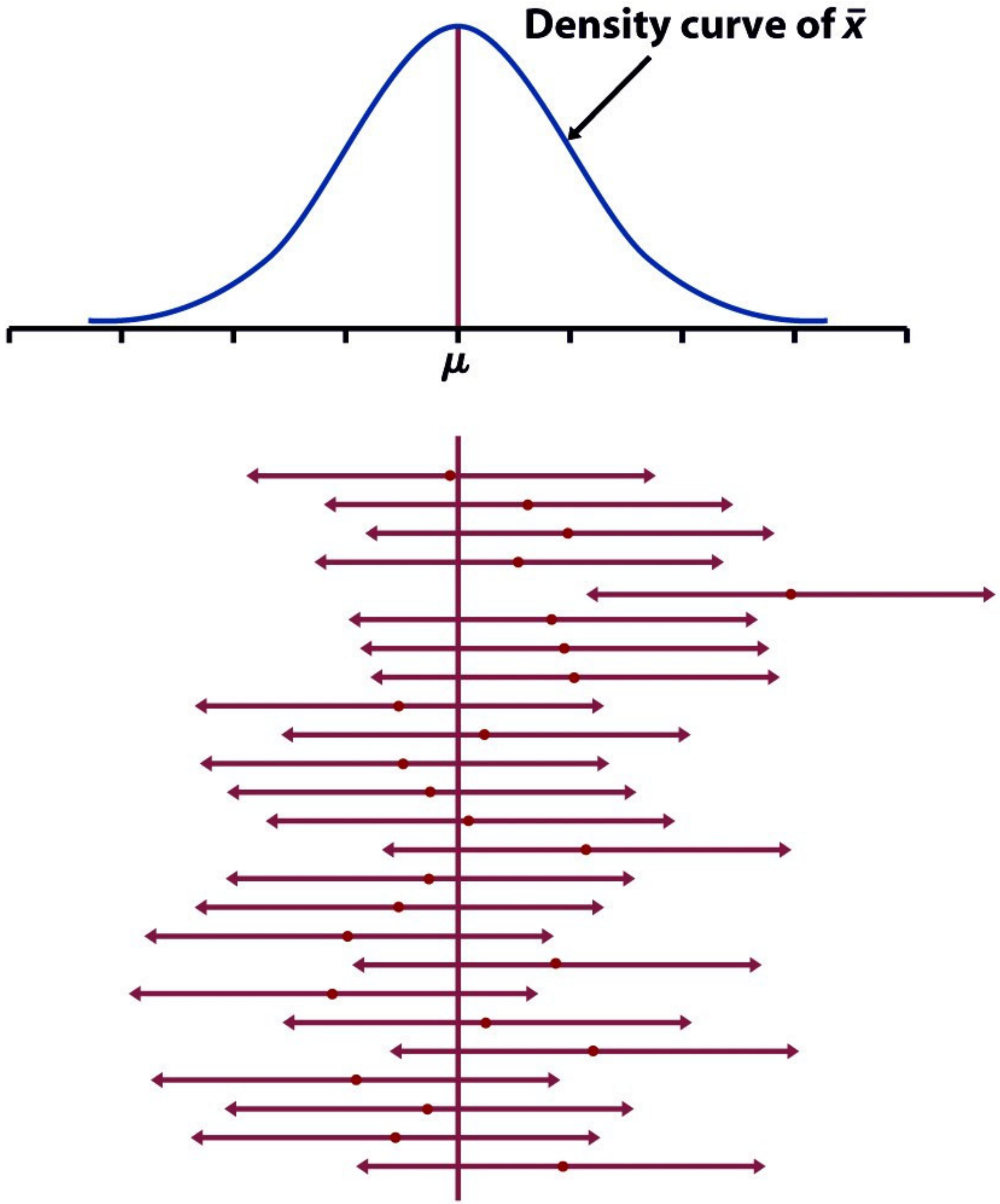


Figure 6-3
Introduction to the Practice of Statistics, Fifth Edition

Hypothesis Testing

- **Very similar to reasoning with confidence intervals, but slightly different perspective:**
  - Confidence interval are used when you want to estimate a population parameter (e.g., μ)
  - Hypothesis testing applies when you want to know the probability of finding some test result (e.g., mean from your sample)

Hypothesis Testing and Significance tests

- CI: good to estimate a population parameter
- Hp testing: is an *inferential* reasoning.
  - It assesses evidences provided by the data in favor of some claim about the population

- SIGNIFICANCE TEST
  - It compares data with *hypotheses* whose truth we want to assess
  - The hp is a claim about a model or a population parameter
  - The results of a significance test are expressed in terms of *probability* (how much are data and Hp similar to each other?

H_o and H_a

- H_0
  - It is the hp of “no difference” or “no effect”. It is the hp we want to test with a significance test.
  - The significance test gives the strength of the evidence against $H_{0.}$ It assess the evidence against H0 in terms of probability.
- Ha
  - Is the case we suspect to be true. It is in any case the hp we would like to proof to be true

Significance test at work

- The significance test is based on a statistic that estimates the parameter that appears in the Hps

- If H0 is true
  - We expect the estimate to take a value near the parameter value specified by H_0
- If H0 is false
  - The values of the estimate are far from the parameter values specified by H_0

Another formulation

- A test of significance finds the probability (p-value) of getting an outcome *as extreme or more than the observed outcome*
  - Extreme means far away from what we expect assuming H_0 as true
- P-value:
  - The smaller the p-value, the stronger the evidences against H_0.

Alfa -level

- ά-level
- It is a level that we consider critical to compare our significance with
  - Typically
    - ά=0.05
      - There will be only 5% of chances that the H0 will be rejected, in case it were true
    - ά=0.01
      - There will be only 1% of chances that the H0 will be rejected, in case it were true

TYPE I AND TYPE II ERRORS

If we reject H_0 (accept H_a) when in fact H_0 is true, this is a **Type I error.** If we accept H_0 (reject H_a) when in fact H_a is true, this is a **Type II error.**

Definition, pg 435
Introduction to the Practice of Statistics, Fifth Edition

		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

Figure 6-17
Introduction to the Practice of Statistics, Fifth Edition

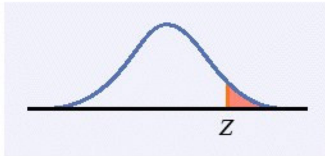
z TEST FOR A POPULATION MEAN

To test the hypothesis H_0: $\mu = \mu_0$ based on an SRS of size n from a population with unknown mean μ and known standard deviation σ, compute the test statistic

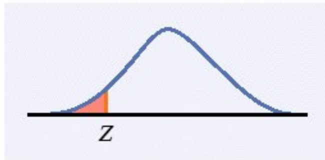
$$z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

In terms of a standard normal random variable Z, the P-value for a test of H_0 against

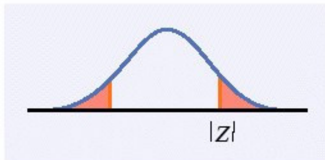
H_a: $\mu > \mu_0$ is $P(Z \geq z)$



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



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



These P-values are exact if the population distribution is normal and are approximately correct for large n in other cases.

Definition, pg 410
Introduction to the Practice of Statistics, Fifth Edition

Hypothesis Testing

- Example book (pp. 409-411):
  - Mean blood pressure in population is 128, $\sigma = 15$
  - Sample of 72 executives with lot of stress
  - Sample average is 126.07, is this ‘normal’?

H_0: $\mu_{executives} = \mu_{population} = 128$
H_a: $\mu_{executives} \neq \mu_{population}$

  - $z = (126.07-128) / (15/ \sqrt{72}) = |1.09|$
  - For z= -1.09 we find in the tables that the probability is 0.1379. Multiplied per 2 = 0.2758

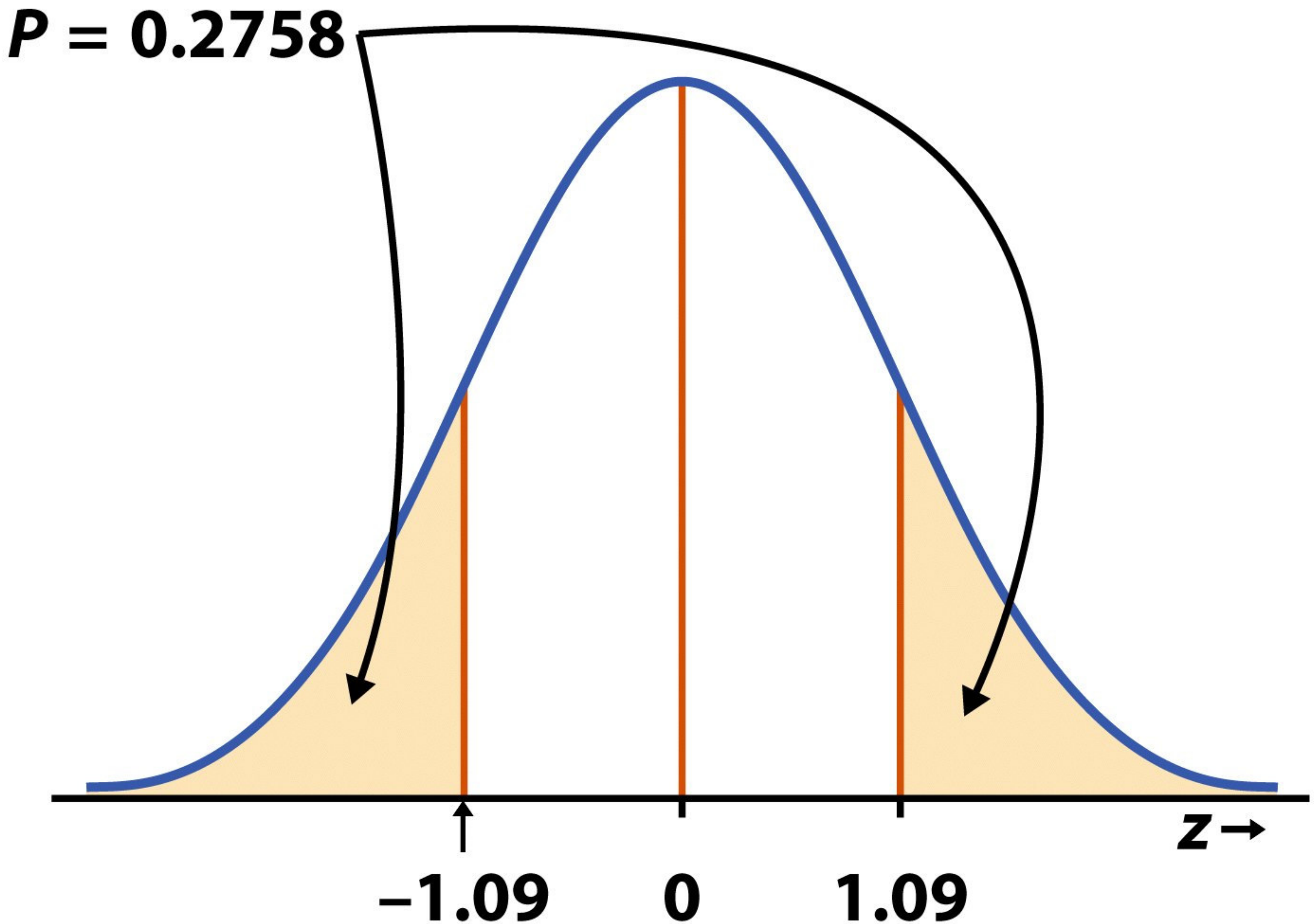


Figure 6-11
Introduction to the Practice of Statistics, Fifth Edition

T-distribution and t-test

T-tests

- Assumption: σ population is UNKNOWN
- This happens most of the time.
  - Usually σ is known when we have standardized tests)
- This hypothesis testing (and associated calculation of confidence intervals) is done with *t-tests*

T-test

- Standardized sample mean, or one-sample z statistic is basis of z-procedures for inference about μ when σ is known:

  z = mean - μ / (σ /√n)
- Most of the time, we only know S, not σ
- Therefore we use S to estimate σ.
- However, when we use **S**/√**n** instead of **σ** /√**n**,

  z does NOT have a **standardnormal** distribution!
- It has a **t distribution**

The z-statistics

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

THE *t* DISTRIBUTIONS

Suppose that an SRS of size n is drawn from an $N(\mu, \sigma)$ population. Then the **one-sample *t* statistic**

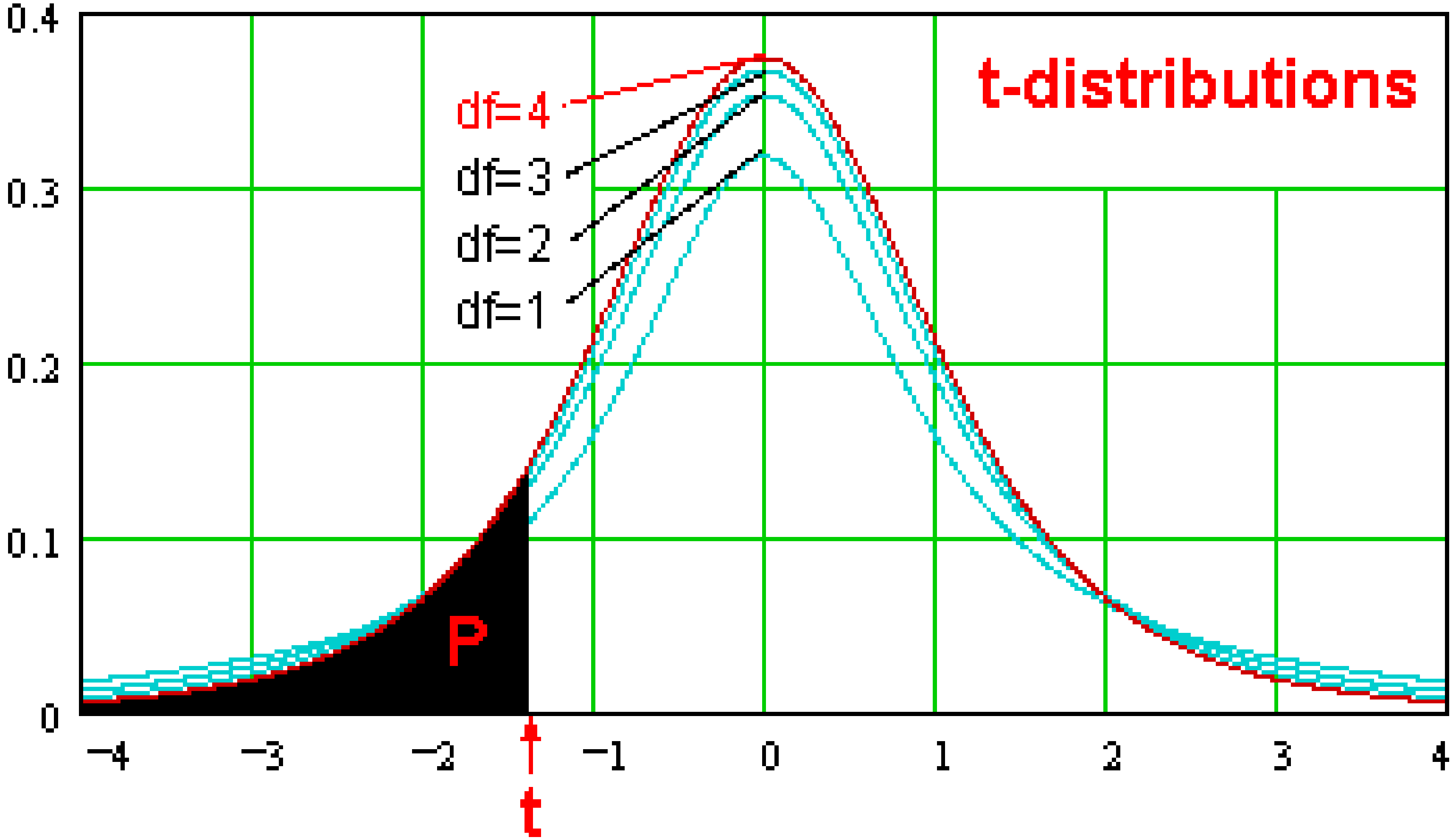
$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

has the ***t* distribution** with $n - 1$ **degrees of freedom.**

Definition, pg 451
Introduction to the Practice of Statistics, Fifth Edition

Degrees of Freedom

- "In general, the degrees of freedom of an estimate is equal to the *number of independent scores* that go into the estimate *minus the number of parameters estimated as intermediate steps* in the estimation of the parameter itself." (Lane)
- "the number of values in the final calculation of a statistic that are free to vary." (Hoffman)
- Difficult concept, but important in *determining the shape of the distribution* you are dealing with. Varies ***roughly*** with number of observations.


t-distributions
df=4
df=3
df=2
df=1
P
t
0.4
0.3
0.2
0.1
0
−4
−3
−2
−1
0
1
2
3
4

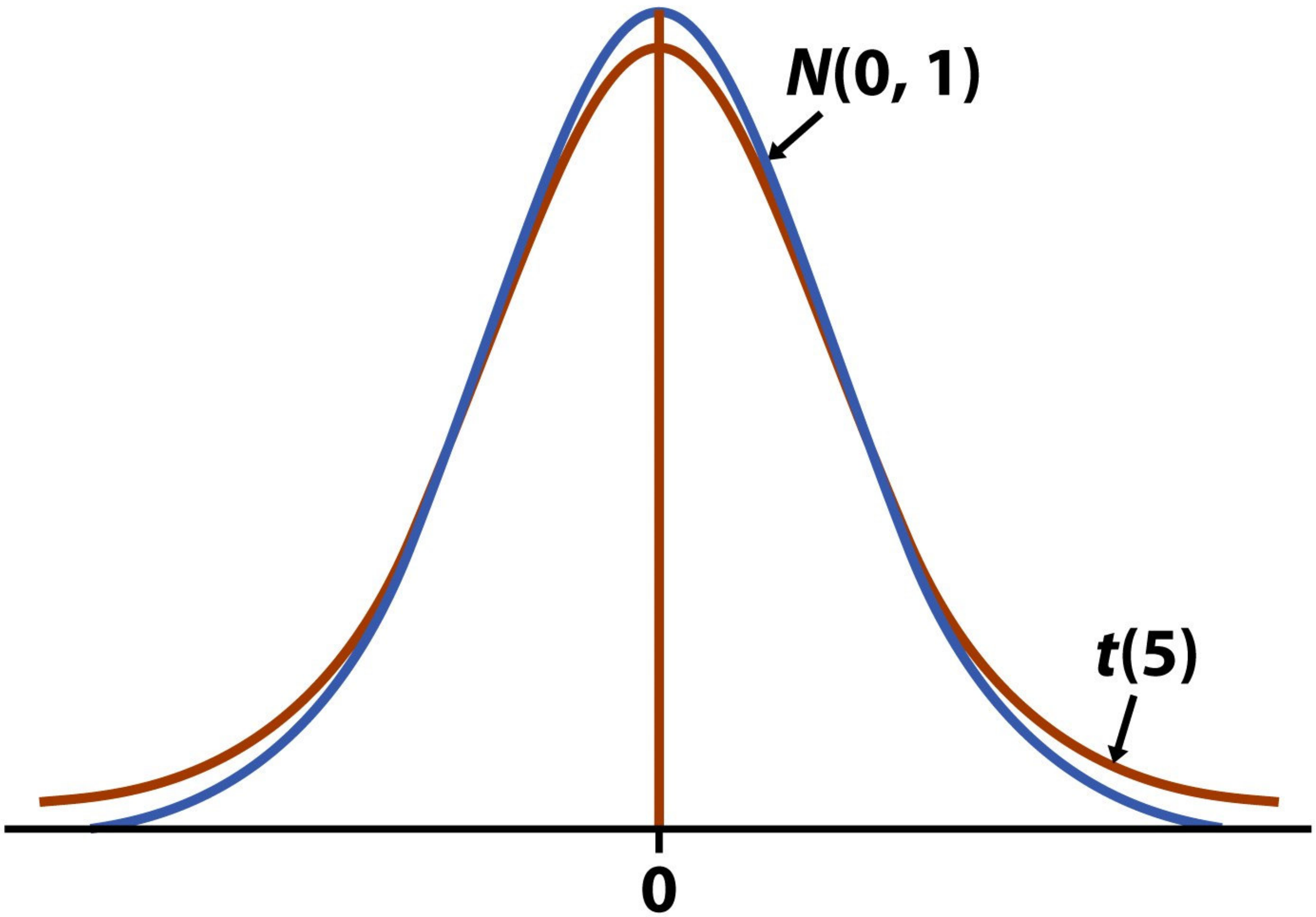


Figure 7-1
Introduction to the Practice of Statistics, Fifth Edition

The t-distribution

- The density curve of t(k) has a similar shape than a normal curve:
  - Symmetric
  - Bell-shaped
- As k>>; t(k) $\rightarrow$ N(0;1)
- Table D shows the critical values for t(k) distributions.

STANDARD ERROR

When the standard deviation of a statistic is estimated from the data, the result is called the **standard error** of the statistic. The standard error of the sample mean is

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Definition, pg 450
Introduction to the Practice of Statistics, Fifth Edition

THE ONE-SAMPLE t CONFIDENCE INTERVAL

Suppose that an SRS of size n is drawn from a population having unknown mean μ. A level C **confidence interval** for μ is

Found in table D

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* is the value for the $t(n-1)$ density curve with area C between $-t^*$ and t^*. The quantity

$$t^* \frac{s}{\sqrt{n}}$$

is the **margin of error.** This interval is exact when the population distribution is normal and is approximately correct for large n in other cases.

Definition, pg 452
Introduction to the Practice of Statistics, Fifth Edition

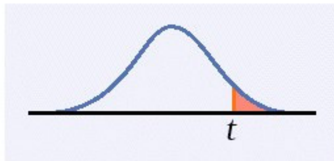
THE ONE-SAMPLE *t* TEST

Suppose that an SRS of size n is drawn from a population having unknown mean μ. To test the hypothesis H_0: $\mu = \mu_0$ based on an SRS of size n, compute the one-sample t statistic

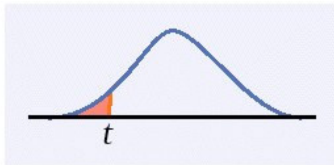
$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

In terms of a random variable T having the $t(n-1)$ distribution, the P-value for a test of H_0 against

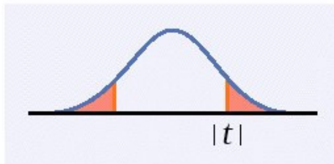
H_a: $\mu > \mu_0$ is $P(T \geq t)$



H_a: $\mu < \mu_0$ is $P(T \leq t)$



H_a: $\mu \neq \mu_0$ is $2P(T \geq |t|)$



These P-values are exact if the population distribution is normal and are approximately correct for large n in other cases.

Definition, pg 454
Introduction to the Practice of Statistics, Fifth Edition

Example

- Vitamin C in a food product: should be 40 mg per 100 grams.
- H_0: $\mu = 40$
- H_a: $\mu \neq 40$
- T-test statistic for

  n=8, mean = 22.50 and Sd=7.19
- What is the p-value?
- Is this significant?

We do the exercise together

- t(7)
- $t = \text{mean} - \mu / (S/\sqrt{n})$
- t=-6.88
- Table D: go to find t=-6.88 with df=7 (n-1)
- $p<0.001 \longrightarrow$ we can reject H_0 and accept H_a

Almost 7 SE (Standard errors) from μ = 40

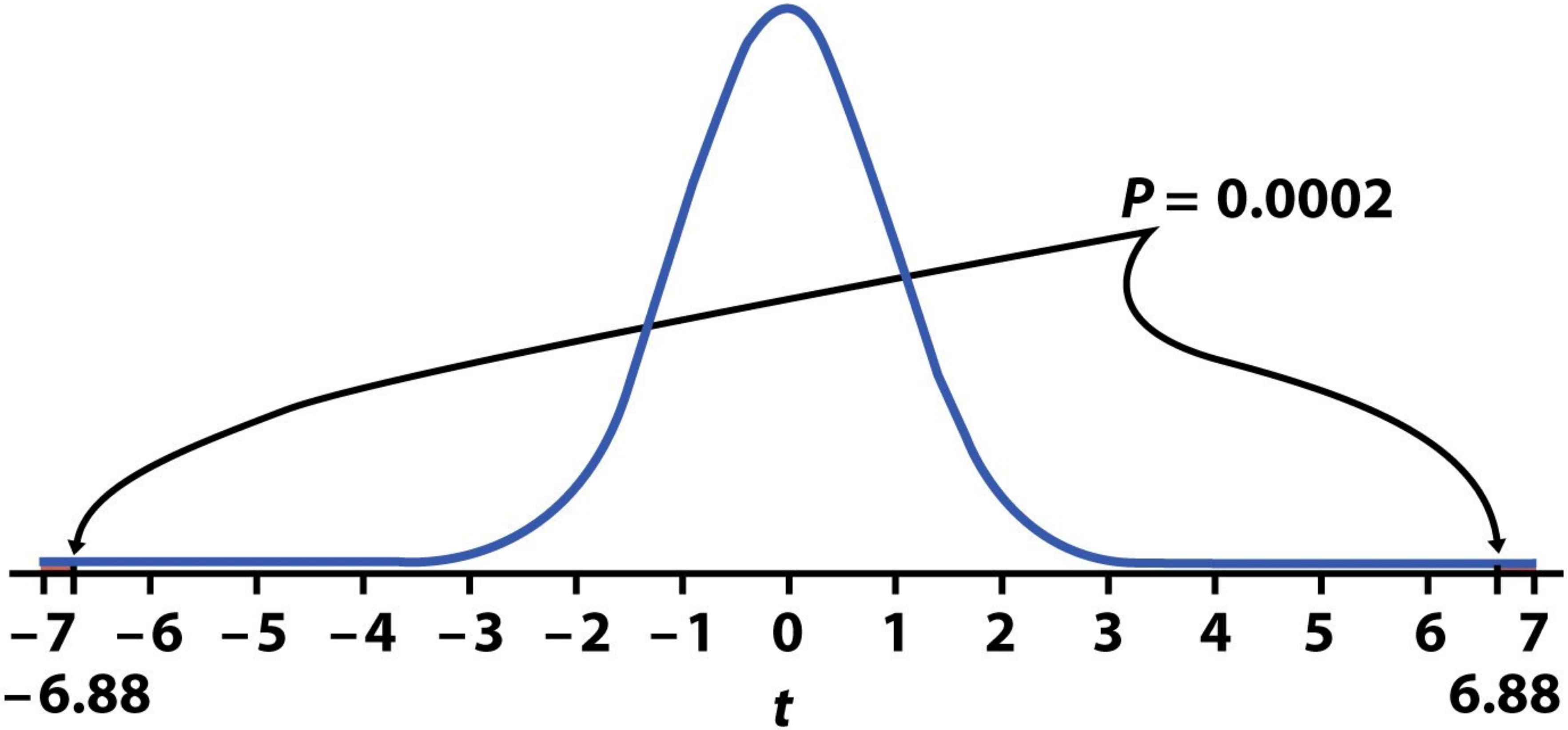


Figure 7-2
Introduction to the Practice of Statistics, Fifth Edition

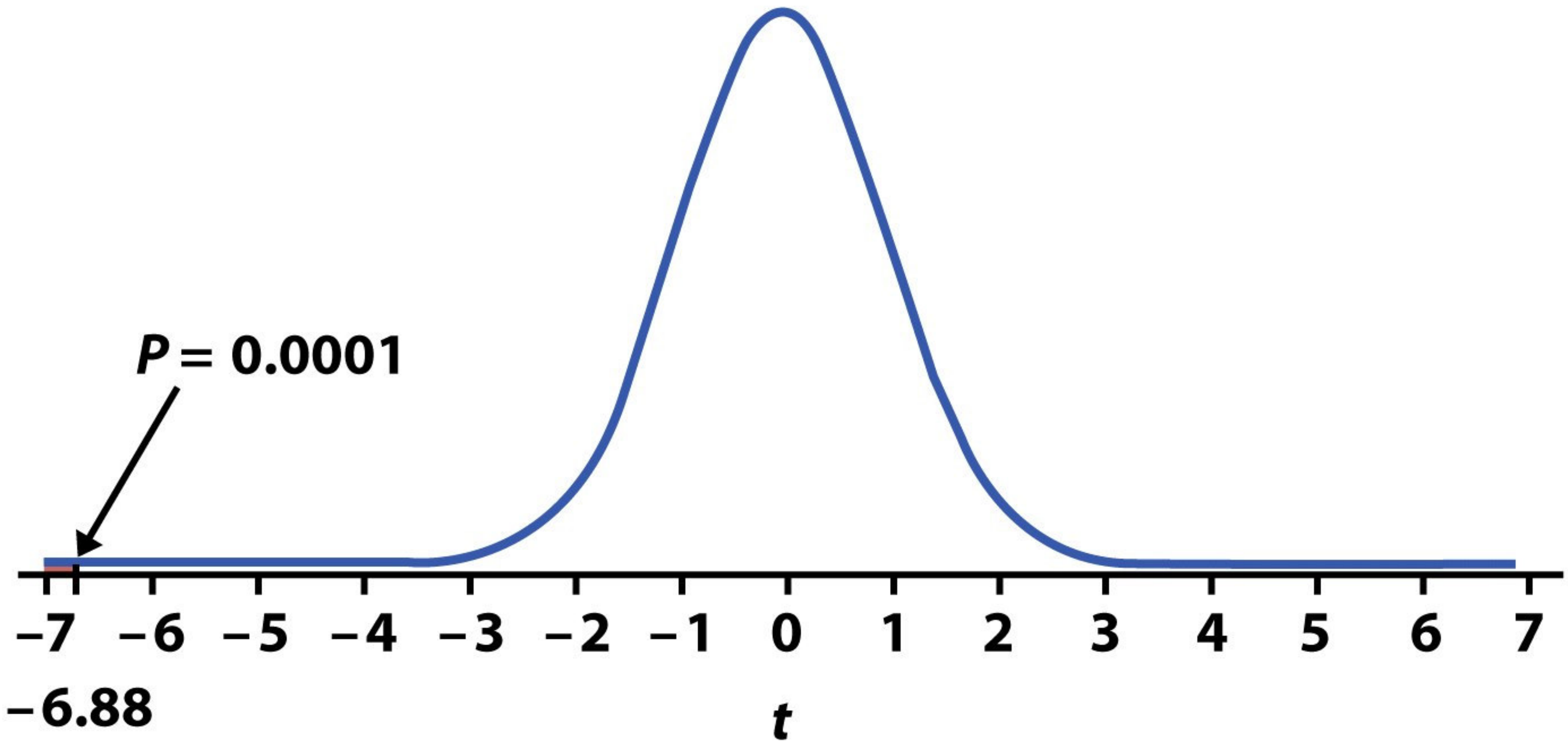


Figure 7-3
Introduction to the Practice of Statistics, Fifth Edition

Paired t-tests

- If you have *matched* observations you can perform a one-sample t-test on the *difference*!
- Observations from the same subject but in different conditions, i.e. number of verb produced by aphasic patients before or after speech therapy.

Example

- Aggressive behaviors of dementia patients on days where there is a full moon
- The observations for both conditions (moon days vs other days) are *matched pairs*. This pair could consist of two subjects, but here, the two observations (moon vs no moon) are from one and the same experimental unit, namely the patient.
- We apply the t-statistics to the differences. We want to see if there is a difference!

TABLE 7.2

Aggressive behaviors of dementia patients

Patient	Moon days	Other days	Difference	Patient	Moon days	Other days	Difference
1	3.33	0.27	3.06	9	6.00	1.59	4.41
2	3.67	0.59	3.08	10	4.33	0.60	3.73
3	2.67	0.32	2.35	11	3.33	0.65	2.68
4	3.33	0.19	3.14	12	0.67	0.69	−0.02
5	3.33	1.26	2.07	13	1.33	1.26	0.07
6	3.67	0.11	3.56	14	0.33	0.23	0.10
7	4.67	0.30	4.37	15	2.00	0.38	1.62
8	2.67	0.40	2.27				

Table 7-2
Introduction to the Practice of Statistics, Fifth Edition

- Mean difference = 2.433; Sd = 1.460
- t(14)

Example

- H_o? H_a?

  $H_o: \mu_{diff} = 0$

  $H_o: \mu_{diff} > 0$

- Calculate the t-statistic?
- What is the p-value?
- Why one-sided testing?

NOTE: in SPSS will calculate the difference for you in the *Paired-Samples T-test*

ROBUST PROCEDURES

A statistical inference procedure is called **robust** if the probability calculations required are insensitive to violations of the assumptions made.

Definition, pg 462
Introduction to the Practice of Statistics, Fifth Edition

Robustness for t-tests

- Number < 15: use t if distribution appears to be normal and there are **no outliers / no strong skew**
- Number ≥ 15: use t if there are **no outliers / no strong skew**
- Number ≥ 40: use t *even when there is a strong skew*

Robustness

- **Note**: if number < 15 and not normal or many outliers / strongly skewed:
  - Transformations
  - Non-parametric tests

TWO-SAMPLE PROBLEMS

- The goal of inference is to compare the responses in two groups.
- Each group is considered to be a sample from a distinct population.
- The responses in each group are independent of those in the other group.

Definition, pg 485
Introduction to the Practice of Statistics, Fifth Edition

- Note 1: regard experimental and control group as coming from distinct populations, even if they *originate* from the same population through random sampling
- Note 2: independence of observations

Variance of *difference*

- We need to know the standard deviation of the *difference* between the two means ($mean_1 - mean_2$)
- Mathematically we can derive that the *variance* of the difference ($mean_1 - mean_2$) is:

  $\sigma_1^2 / n_1 \quad + \sigma_2^2 / n_2$
- So the standard deviation, which is the square root of the variance, is the square root of this formula

TWO-SAMPLE *z* STATISTIC

Suppose that $\bar{x}_1$ is the mean of an SRS of size n_1 drawn from an $N(\mu_1, \sigma_1)$ population and that $\bar{x}_2$ is the mean of an independent SRS of size n_2 drawn from an $N(\mu_2, \sigma_2)$ population. Then the **two-sample *z* statistic**

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

has the standard normal $N(0, 1)$ sampling distribution.

Definition, pg 488
Introduction to the Practice of Statistics, Fifth Edition

THE TWO-SAMPLE t SIGNIFICANCE TEST

Suppose that an SRS of size n_1 is drawn from a normal population with unknown mean μ_1 and that an independent SRS of size n_2 is drawn from another normal population with unknown mean μ_2. To test the hypothesis H_0: $\mu_1 = \mu_2$, compute the **two-sample *t* statistic**

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

and use P-values or critical values for the $t(k)$ distribution, where the degrees of freedom k are either approximated by software or are the smaller of $n_1 - 1$ and $n_2 - 1$.

Definition, pg 489

Introduction to the Practice of Statistics, Fifth Edition

back

Two sample t-statistic

- Estimates σ_1 with Sd_1 and σ_2 with Sd_2
- No z-distribution
- No **t-distribution**!
- But *approximation* of t-distribution t(k) where k is the *approximation* of the degrees of freedom (done by software)

Example

- One class of 21 schoolkids get extra reading activities
- One class of 23 kids (control group) does not get extra activities
- P. 490 gives back-to-back stemplot
- plus normal quantile plots

TABLE 7.4

DRP scores for third-graders

Treatment group				Control group			
24	61	59	46	42	33	46	37
43	44	52	43	43	41	10	42
58	67	62	57	55	19	17	55
71	49	54		26	54	60	28
43	53	57		62	20	53	48
49	56	33		37	85	42	

Table 7-4
Introduction to the Practice of Statistics, Fifth Edition

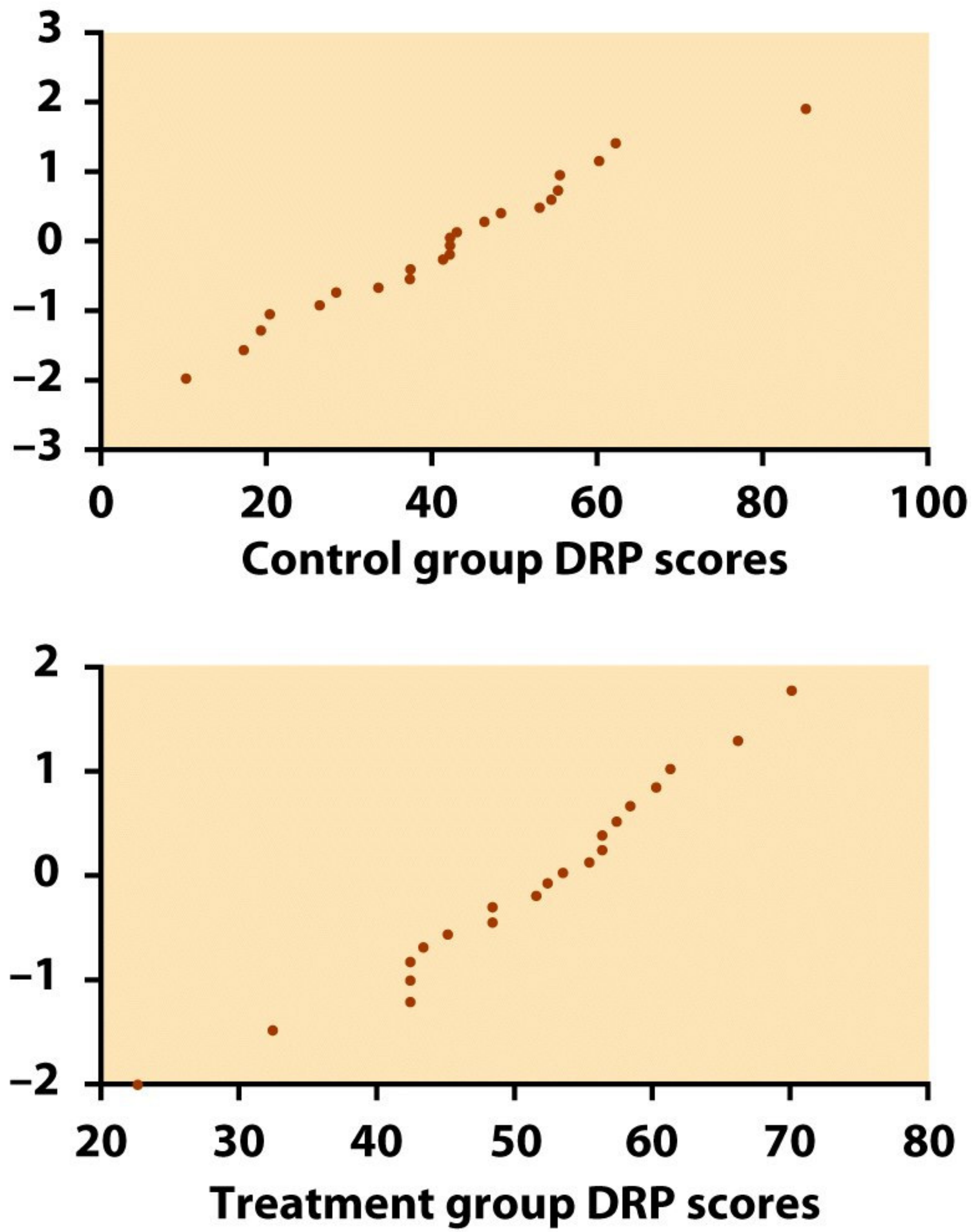


Figure 7-13
Introduction to the Practice of Statistics, Fifth Edition

Example

- Descriptive statistics:
  - N (treatment) = 21; mean = 51.48; Sd = 11.01
  - N (control) = 23; mean = 41.52; Sd = 17.15

  - Nullhypothesis: $\mu_{treat} = \mu_{control}$
  - Alternative hypothesis: $\mu_{treat} > \mu_{control}$

- Two-sample t-statistic

Levene's test

- Tests whether to reject the null hypothesis of equal variances

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std Diff
DEV	Equal variances assumed	,236	,630	9,909	42	,000	18,5917	1
	Equal variances not assumed			9,792	38,258	,000	18,5917	1

Summary

- t-test is performed when σ is unknown (and estimated with Sd)
- One-sample t-test has t-distribution with df = n-1
- Matched Pairs t-test can be performed with one-sample t-test
- Two-sample t-test has an approximate t-distribution; df is calculated by software