Statistical and Mathematical Methods for Data Analysis

Dr. Faisal Bukhari

Punjab University College of Information Technology (PUCIT)

Textbook

- **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

References

Readings for these lecture notes:

- Probability & Statistics for Engineers & Scientists, Ninth edition, Ronald E. Walpole, Raymond H. Myer

- Elementary Statistics, Tenth Edition, Mario F. Triola

These notes contain material from the above resources.

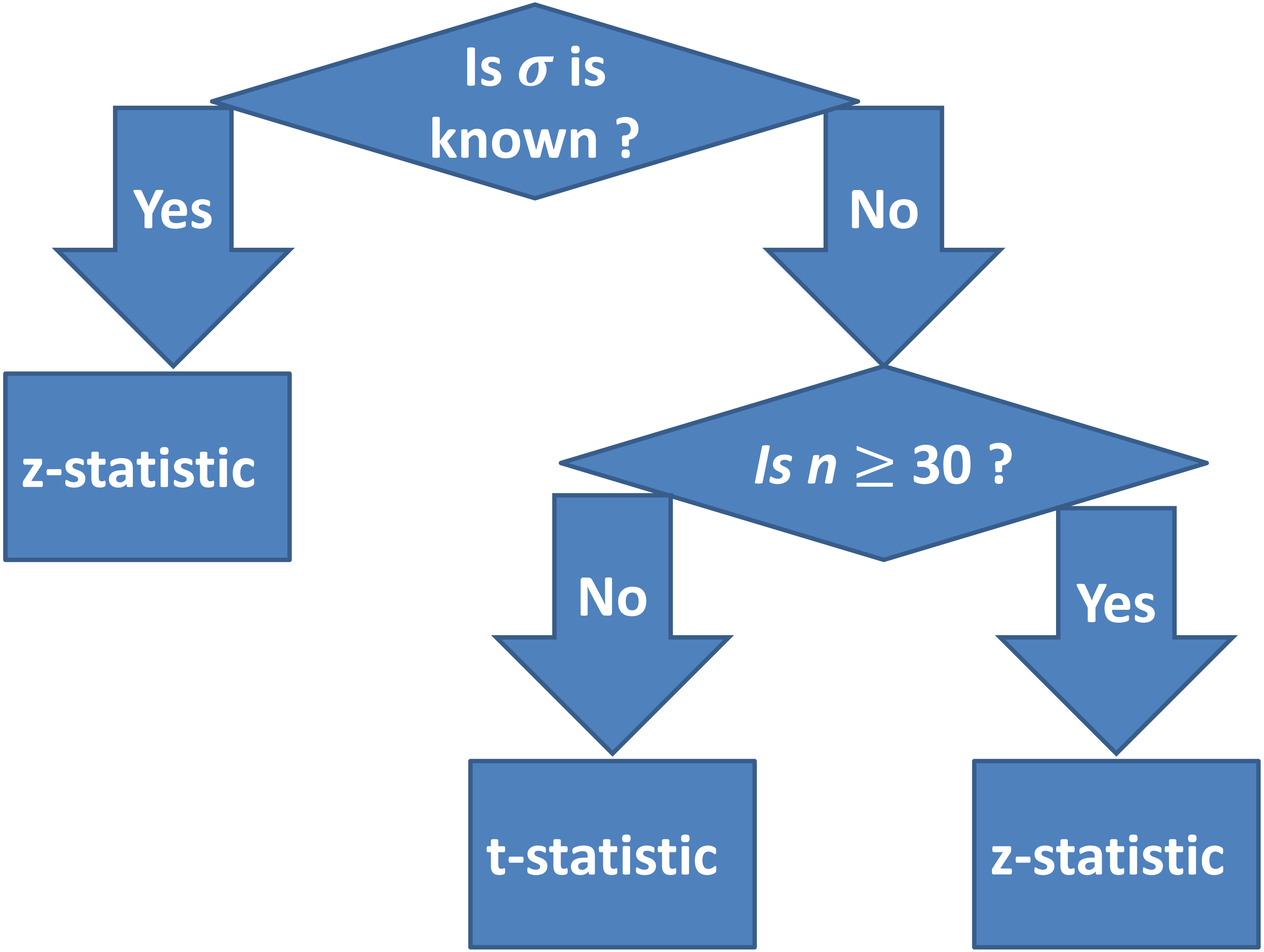
Is σ is known ?

Yes

If either the population is normally distributed or $n \geq 30$, then use the use the standard normal distribution or Z-test

No

If either the population is normally distributed or $n \geq 30$, then use the *t*-distribution or t-test


Is σ is known ?
Yes
No
z-statistic
Is $n \geq 30$?
No
Yes
t-statistic
z-statistic

When both ***n* < 30** and the **population is *not* normally distributed**, we *cannot* use the **standard normal distribution** or the ***t*-distribution**.

The two main activities of inferential statistics are using sample data to

(1) *estimate* a **population parameter**, and

(2) test a **hypothesis or claim** about a **population parameter**

Hypothesis

- In statistics, a **hypothesis** is a **claim** or **statement** about a **property of a population**.

- A **hypothesis test** (or **test of significance**) is a **standard procedure** for **testing a claim** about a property of a population.

Examples of Hypotheses

- **Business** A newspaper headline makes the claim that most workers get their jobs through **networking**.

- **Medicine** Medical researchers claim that the **mean body temperature** of healthy adults is not equal to 98.6°F.

- **Aircraft Safety** The Federal Aviation Administration claims that the **mean weight** of an airline passenger (with carry on baggage) is greater than the **185 lb** that it was 20 years ago.

Type I Error : Rejection of the null hypothesis when it is true is called a **type 1 error**.

Type II error : Nonrejection of the null hypothesis when it is false is called a **type II error**.

	H_0 is true	**H_0 is false**
Do not reject H_0	Correct decision	**Type II error**
Reject H_0	**Type I error**	Correct decision

Verdict	Truth About Defendant	
	Innocent	Guilty
Not Guilty	Justice	Type II error
Guilty	Type I error	Justice

Approach to Hypothesis Testing with Fixed Probability of Type I Error

1. State the **null** and **alternative hypotheses**.
2. Choose a fixed **significance level α**.
3. **Test statistic** to be used it
4. **Calculations**
5. **Critical region**
6. **Conclusion**

Area under the Normal Curve [1]

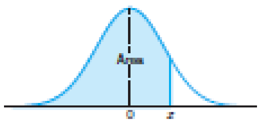


Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Area under the Normal Curve [2]

Table A.3 (continued) Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

H_0	Value of Test Statistic	H_1	Critical Region
$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$; σ known	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$; $v = n - 1$, σ unknown	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; σ_1 and σ_2 known	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$; $v = n_1 + n_2 - 2$, $\sigma_1 = \sigma_2$ but unknown, $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t' = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$; $v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$, $\sigma_1 \neq \sigma_2$ and unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t' < -t_\alpha$ $t' > t_\alpha$ $t' < -t_{\alpha/2}$ or $t' > t_{\alpha/2}$
$\mu_D = d_0$ paired observations	$t = \dfrac{\bar{d} - d_0}{s_d/\sqrt{n}}$; $v = n - 1$	$\mu_D < d_0$ $\mu_D > d_0$ $\mu_D \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

A null hypothesis H_0 vs the alternative hypothesis H_1

A **null hypothesis** H_0 **is** a statistical hypothesis that contains a statement of equality, such as $\leq, \geq$ $\boldsymbol{or}$ $=$

The **alternative hypothesis** H_1 is the complement of the null hypothesis. It is a statement that must be true if H_0 is false and it contains a statement of **strict inequality**, such as **>, <, or** $\neq$

A null hypothesis H_0 vs the alternative hypothesis H_1

To write the null and alternative hypotheses, **translate the claim** made about the **population parameter** from a **verbal statement** to a **mathematical statement**.

Then, write its complement. For instance, if the claim value is k and the population parameter is μ, then some possible pairs of null and alternative hypotheses are

$$\begin{cases} \mathrm{H_0}: \mu \leq k \\ \mathrm{H_1}: \mu > k \end{cases} \qquad \begin{cases} \mathrm{H_0}: \mu \geq k \\ \mathrm{H_1}: \mu < k \end{cases} \qquad \begin{cases} \mathrm{H_0}: \mu = k \\ \mathrm{H_1}: \mu \neq k \end{cases}$$

A null hypothesis H_0 vs the alternative hypothesis H_1

Regardless of which of the **three pairs of hypotheses** you use, you always assume $\mu = k$ and examine the **sampling distribution** on the basis of this assumption.

Within this sampling distribution, you will determine whether or not a **sample statistic** is unusual.

Null Hypothesis

The **null hypothesis** (denoted by H_0) is a statement that the value of a **population parameter** (such as proportion, mean, or standard deviation) is equal to some **claimed value**. Here are some typical null hypotheses of the type:

H_0: $p = 0.5$ **H_0: $\mu = 98.6$** **H_0: $\sigma = 15$**

- We test the null hypothesis directly in the sense that we **assume it is true** and reach a conclusion to either reject H_0 or fail to reject H_0.

Alternative Hypothesis

- The **alternative hypothesis** (denoted by H_1 or H_a or H_A) is the statement that the parameter has a value that somehow differs from the null hypothesis. The symbolic form of the alternative hypothesis must use one of these symbols:

  $<$ or $>$ or $\neq$.

Alternative Hypothesis

Here are nine different examples of alternative hypotheses involving proportions, means, and standard deviations:

- **Means:** H_1: **μ** > 98.6 H_1: < 98.6 H_1: ≠ 98.6

- **Proportions:** H_1:p > 0.5 H_1 :p< 0.5 H_1:p≠ 0.5

- **Standard Deviations:H_1**: $\sigma > 15$ H_1: $\sigma < 15$ H_1: $\sigma \neq 15$

Note About Always Using the Equal Symbol in H_0:

- A few textbooks use the symbols $\geq$ and $\leq$ in the null hypothesis H_0, but most professional journals use only the equal symbol for equality.

- We conduct the hypothesis test by assuming that the proportion, mean, or standard deviation is *equal to* some specified value so that we can work with a single distribution having a specific value.

Note About Forming Your Own Claims (Hypotheses):

- If you are conducting a study and want to use a hypothesis test to *support* your claim, **the claim must be worded so that it becomes the alternative hypothesis** (and can be expressed using only the symbols **$<$ or $>$ or $\neq$**).

- You can never support a claim that some parameter is *equal to* some specified value

- For example, if you have developed a gender-selection method that increases the likelihood of a girl, state your claim as ***p* > 0.5** so that your claim can be supported.
- (In this context of trying to support the goal of the research, the **alternative hypothesis** is sometimes referred to as the ***research hypothesis***.)
- You will assume for the purpose of the test that ***p* = 0.5**, but you hope that $p = 0.5$ gets rejected so that ***p* > 0.5** is supported.

The Null and Alternative Hypotheses

The structure of hypothesis testing will be formulated with the use of the term **null hypothesis**, which refers to any hypothesis we wish to test and is denoted by H_0. The rejection of H_0 leads to the acceptance of an **alternative hypothesis**, denoted by H_1.

The alternative hypothesis ***H_1* usually represents the *question to be answered or the theory to be tested,*** and thus its specification is crucial. The null hypothesis H_0 *nullifies or opposes* H_1 and is often the logical complement to H_1.

Identifying H_0 and H_1

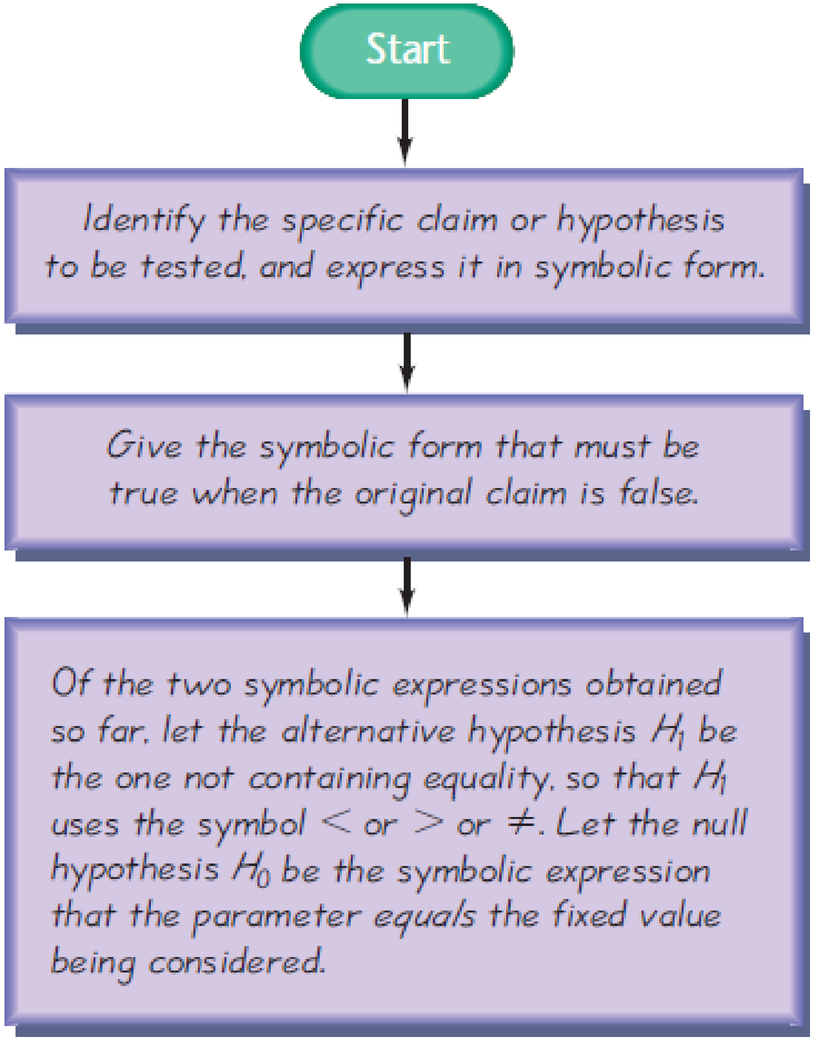


Figure 1

Identifying H_0 and H_1

- Note that the **original statement** could become the **null hypothesis**, it could become the **alternative hypothesis**, or **it might not correspond exactly** to either the **null hypothesis** or the **alternative hypothesis.**

- For example, we sometimes test the validity of someone else's claim, such as the claim of the **Coca-Cola Bottling Company** that **"the mean amount of Coke in cans is at least 12 oz."** That claim can be expressed in symbols as $\mu \geq 12$.

- In Figure 1 we see that if that original claim is false, then $\mu < 12$. The **alternative hypothesis** becomes $\mu < 12$, but the **null hypothesis** is $\mu = 12$.

- We will be able to address the **original claim** after determining whether there is sufficient evidence to reject the null hypothesis of $\mu = 12$.

EXAMPLE Identifying the Null and Alternative Hypotheses Refer to Figure 8-2 and use the given claims to express the corresponding null

and alternative hypotheses in symbolic form.

a. The **proportion** of workers who get jobs through networking is **greater than 0.5**.

b. The **mean weight** of airline passengers with carry-on baggage is at **most 195 lb** (the current figure used by the Federal Aviation Administration).

c. The standard deviation of IQ scores of actors is equal to **15**.

SOLUTION See Figure 1, which shows the three-step procedure.

- In Step 1 of Figure 1, we express the given **claim as $p > 0.5$**. In Step 2 we see that if **$p > 0.5$** is **false**, then **$p \leq 0.5$** must be true. In Step 3, we see that the expression **$p > 0.5$** does not contain equality, so we let the **alternative hypothesis H_1** be **$p > 0.5$**, and we let H_0 be $p = 0.5$.

- In Step 1 of Figure 1, we express **"a mean of at most 195 lb"** in symbols as $\mu \leq 195$ In Step 2 we see that if $\mu \leq 195$ is false, then $\mu > 195$ must be true. In Step 3, we see that the expression $\mu > 195$ does not contain equality, so we let the alternative hypothesis H_1: $\mu > 195$ be and we let H_0 be $\mu = 195$

c. In Step 1 of Figure 1, we express the given claim as **σ = 15** In Step 2 we see that if $\sigma = 15$ is false, then $\sigma \neq 15$ must be true. In Step 3, we let the alternative hypothesis H_1 be $\sigma \neq 15$, and we let **H_0 be $\sigma = 15$** .

Single Sample: Tests Concerning a Single Mean

Example: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Solution:

n =100 (sample size)

$\bar{x} = 71.8$ (sample mean)

σ = 8.9 (population standard deviation)

α = 0.05 (level of significance)

1. **We state our hypothesis as**:

H_0: μ = 70 years

H_1: μ > 70 years (one sided test)

2. **The level of significance is set** $\alpha = 0.05$.

3. **Test statistic to be used is**

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

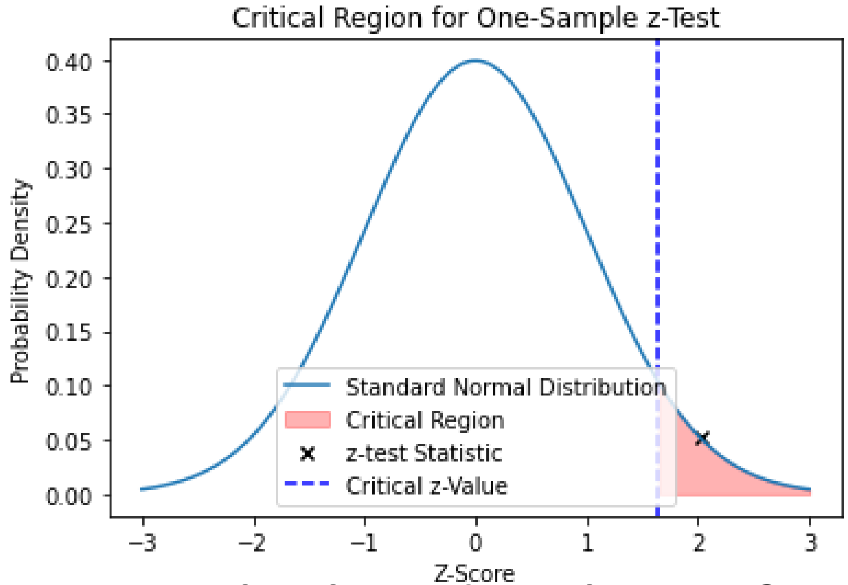
4. **Calculations:**

$Z_{cal} = \;\; 2.02$

5. **Critical region:**

$Z_{cal} > Z_{tab}$

Where $Z_{tab} = Z_{\alpha} = Z_{0.05} = 1.6405$ $\because 1 - \alpha = 1 - 0.05 = 0.95$



6. **Conclusion:** Since calculated value of Z is greater than the tabulate value of Z, so we are unable to accept H_O

```python
import numpy as np
from scipy.stats import norm

# Given data
sample_mean = 71.8
population_std_dev = 8.9
sample_size = 100
population_mean_hypothesis = 70
significance_level = 0.05

# Calculate the standard error of the mean
standard_error = population_std_dev / np.sqrt(sample_size)

# Calculate the Z-statistic
z_statistic = (sample_mean - population_mean_hypothesis) /
standard_error

# Find the critical Z-value for a one-tailed test
critical_z_value = norm.ppf(1 - significance_level)
```

```python
# Compare Z-statistic to critical Z-value
print("Z-statistic:", round(z_statistic, 4))
print("Critical Z-value:", round(critical_z_value, 4))

if z_statistic > critical_z_value:
    print("Reject the null hypothesis. There is enough evidence to suggest that the mean life span today is greater than 70 years.")
else:
    print("Fail to reject the null hypothesis. There is not enough evidence to suggest that the mean life span today is greater than 70 years.")
```

Output

```
Z-statistic: 2.0225
Critical Z-value: 1.6449
Reject the null hypothesis. There is enough evidence to suggest
that the mean life span today is greater than 70 years.
```

Example: A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of **8 kilograms** with a **standard deviation of 0.5** kilogram. Test the hypothesis that $\mu = 8$ kilograms against the alternative that $\mu \neq 8$ kilograms if a random sample of **50 lines** is tested and found to have a mean breaking strength of **7.8 kilograms**. Use a 0.01 level of significance.

Solution:

$\mu = 8$ (Population mean)

$n = 50$ (Sample size)

$\sigma = 0.5$ (Population standard deviation)

$\bar{x} = 7.8$ (Sample mean)

$\alpha = 0.01$ (Level of significance)

1. **We state our hypothesis as**:

H_0: $\mu = 8$

H_1: $\mu \neq 8$ (Two sided test)

2. **The level of significance is set** $\alpha = 0.01$.

3. **Test statistic to be used is**

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

4. **Calculations:**

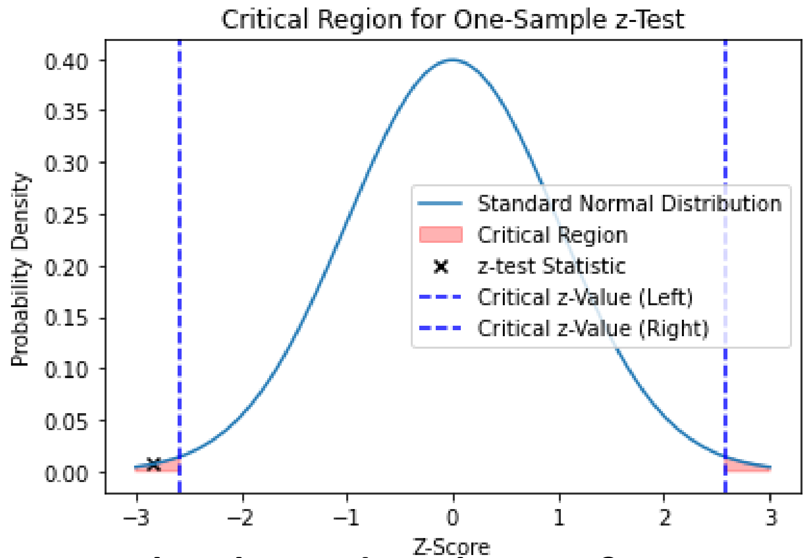
$$Z_{cal} = \frac{7.8 - 8}{0.5/\sqrt{50}} = -2.83.$$

5. **Critical region:**

$|Z_{cal}| > Z_{tab}$

Where $Z_{tab} = Z_{\alpha/2} = Z_{0.005} = 2.575$

$\because 1 - \alpha/2 = 1 - 0.01/2 = 0.995$



6. **Conclusion:** Since calculated value of Z is greater than the tabulate value of Z, so we are unable to accept H_O

```python
import numpy as np
from scipy.stats import norm

# Given data
sample_mean = 7.8
population_std_dev = 0.5
sample_size = 50
population_mean_hypothesis = 8
significance_level = 0.01

# Calculate the standard error of the mean
standard_error = population_std_dev / np.sqrt(sample_size)

# Calculate the Z-statistic
z_statistic = (sample_mean - population_mean_hypothesis) / 
standard_error
```

```python
# Find the critical Z-values for a two-tailed test
critical_z_value_positive = norm.ppf(1 - significance_level / 2)
critical_z_value_negative = -critical_z_value_positive

# Compare Z-statistic to critical Z-values
print("Z-statistic:", round(z_statistic, 4))
print("Critical Z-values:", round(critical_z_value_negative, 4), round(critical_z_value_positive, 4))

if z_statistic < critical_z_value_negative or z_statistic > critical_z_value_positive:
    print("Reject the null hypothesis. There is enough evidence to suggest that the mean breaking strength is different from 8 kilograms.")
else:
    print("Fail to reject the null hypothesis. There is not enough evidence to suggest that the mean breaking strength is different from 8 kilograms.")
```

Output

```
Z-statistic: -2.8284
Critical Z-values: -2.5758 2.5758
Reject the null hypothesis. There is enough evidence to suggest
that the mean breaking strength is different from 8 kilograms.
```

Example 10.5: The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of **46** kilowatt hours per year. If a random sample of **12** homes included in a planned study indicates that vacuum cleaners use an average of **42** kilowatt hours per year with a standard deviation of **11.9** kilowatt hours, does this suggest at the **0.05** level of significance that vacuum cleaners use, on average, less than **46** kilowatt hours annually? Assume the population of kilowatt hours to be normal.

Solution

$\mu = 46$ (Population mean)

$n = 12$ (Sample size)

$s = 11.9$ (Sample standard deviation)

$\bar{x} = 42$ (Sample mean)

$\alpha = 0.05$ (Level of significance)

1. **We state our hypothesis as**:

$H_0: \mu = 46$

$H_1: \mu < 46$ (One tailed test)

2. **The level of significance is set** $\alpha = 0.05$.

3. **Test statistic to be used is**

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

4. **Calculations:**

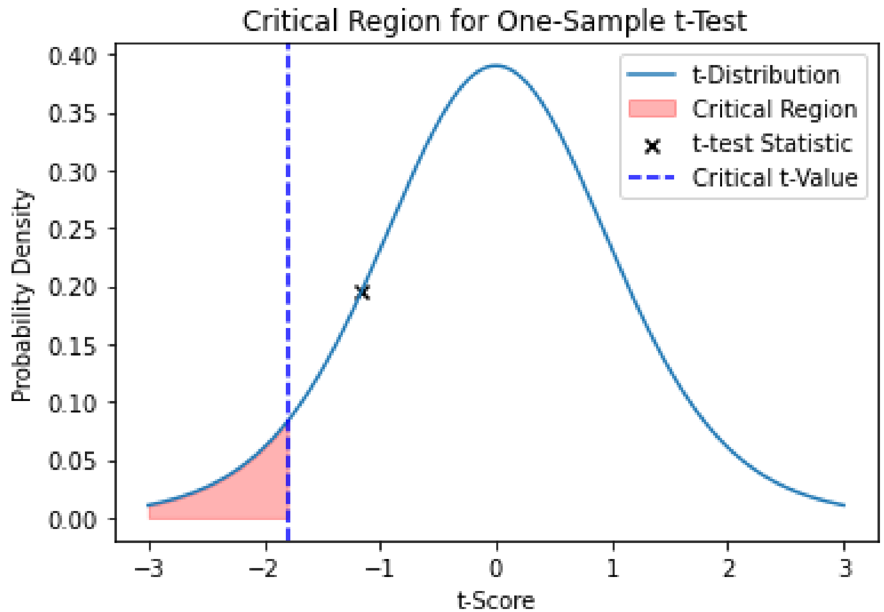
$$t_{cal} = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16$$

5. **Critical region:**

$t_{cal} < t_{tab}$

Where $-t_{tab} = -t_{(\alpha,\, n-1)} = -t_{(0.05,\, 11)} = -1.796$

$-1.16 < -1.796$ (False)



6. **Conclusion:** Since calculated value of t_{cal} is greater than the tabulate value of t, so we accept H_O

```python
import numpy as np
from scipy import stats

# Given data
sample_mean = 42                      # Average kilowatt hours
population_std_dev = 11.9             # Population SD
sample_size = 12                      # Sample size
population_mean_hypothesis = 46       # Ho

# Calculate the standard error of the mean (SEM)
standard_error = population_std_dev / np.sqrt(sample_size)

# Calculate the t-statistic
t_statistic = (sample_mean - population_mean_hypothesis) / standard_error

# Degrees of freedom
degrees_of_freedom = sample_size - 1

# Set the significance level
alpha = 0.05
```

```python
# Calculate the critical t-value for a one-tailed test
critical_t_value = stats.t.ppf(alpha, degrees_of_freedom)

# Round the results to 4 decimal places
t_statistic = round(t_statistic, 4)
critical_t_value = round(critical_t_value, 4)

# Print the results
print(f"t-statistic: {t_statistic}")
print(f"Critical t-value: {critical_t_value}")

# Compare the t-statistic with the critical t-value

if t_statistic < critical_t_value:
    print("Reject the null hypothesis. There is enough evidence to suggest that vacuum cleaners use, on average, less than 46 kilowatt hours annually.")
else:
    print("Fail to reject the null hypothesis. There is not enough evidence to suggest that vacuum cleaners use, on average, less than 46 kilowatt hours annually.")
```

Output:

```
t-statistic: -1.1644
Critical t-value: -1.7959
Fail to reject the null hypothesis. There is not enough evidence to suggest that vacuum cleaners use, on average, less than 46 kilowatt hours annually.
```

Table A.4 Critical Values of the t-Distribution

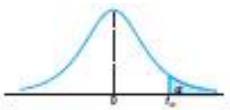


Table A.4 Critical Values of the t-Distribution

	α						
v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Table A.4 (continued) Critical Values of the t-Distribution

v	0.02	0.015	0.01	0.0075	0.005	0.0025	0.0005
1	15.894	21.205	31.821	42.433	63.656	127.321	636.578
2	4.849	5.643	6.965	8.073	9.925	14.089	31.600
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.689
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.660
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.290