Weeks 7-8

Comparison of Population Means

Dr. Esam Mahdi
Department of Mathematics, Statistics and Physics
Qatar University

Learning Objectives

- Test the hypothesis of population mean using one-sample t test.
- Compare the two population means (using parametric t-test):
  - Case 1: Two dependent populations (pair-matched data).
  - Case 2: Two independent populations.
- Compare the two population median (using nonparametric test):
  - Case 1: Pair-matched data using Wilcoxon signed-rank test.
  - Case 2: Two independent populations using Wilcoxon rank-sums test.
- Test whether there are simultaneously differences in the means of several independent groups use the one-way analysis of variance (ANOVA) approach.

Hypothesis Tests about the Population Mean

- When the population is normal and σ is unknown, where the sample size n is small, we will can use t-test statistic

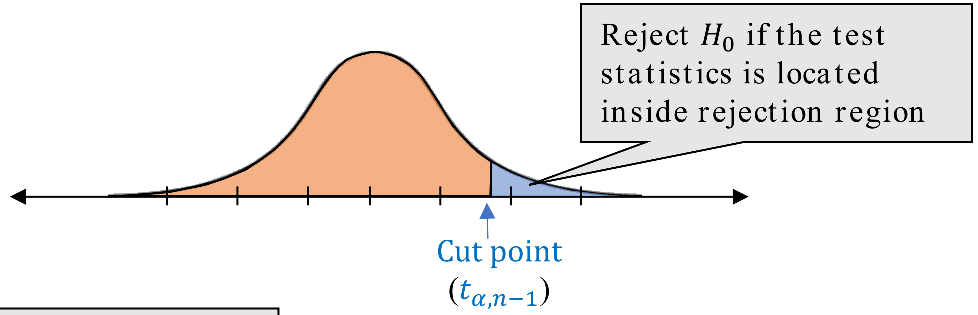
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- The tabular value will be found from (Appendix C) with $(n-1)$ degrees of freedom and significance of level α.

  - One-sided (right tailed) test

$$H_0\colon \mu = \mu_0 \text{ versus } H_a\colon \mu > \mu_0$$

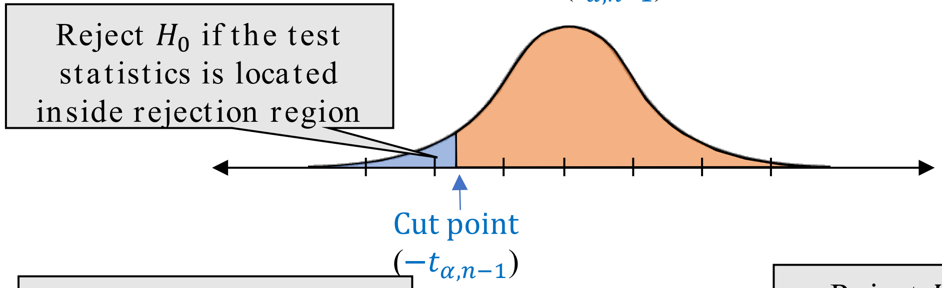
Reject H_0 at the significance of level α if the $\mathrm{t.test} \geq t_{\alpha,n-1}$



  - One-sided (left tailed) test

$$H_0\colon \mu = \mu_0 \text{ versus } H_a\colon \mu < \mu_0$$

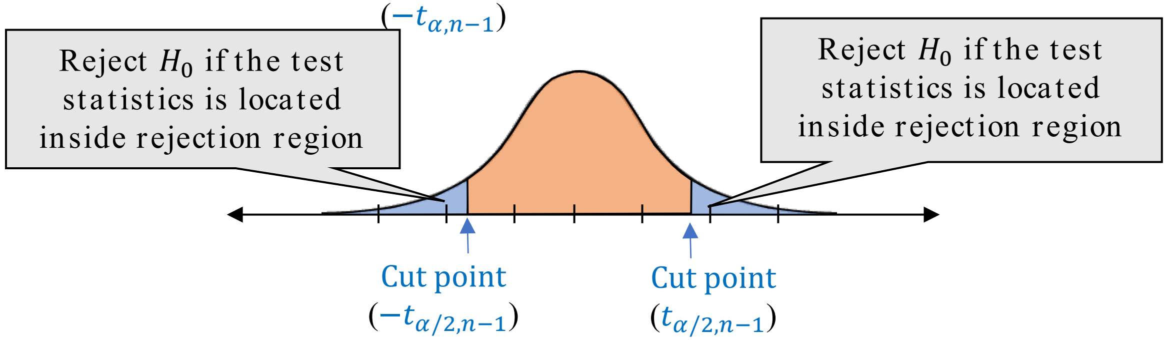
Reject H_0 at the significance of level α if the $\mathrm{t.test} \leq -t_{\alpha,n-1}$



  - Two-sided (two tailed) test

$$H_0\colon \mu = \mu_0 \text{ versus } H_a\colon \mu \neq \mu_0$$

Reject H_0 at the significance of level α if the $\mathrm{t.test} \geq t_{\alpha/2,n-1}$ or $\mathrm{t.test} \leq -t_{\alpha/2,n-1}$



Steps For One-Sample t Test

1. Decide whether a one- or a two-sided test is appropriate; this decision depends on the research question.
2. Choose a level of significance; a common choice is $\alpha = 0.05$.
3. Calculate the t statistic (Note: we can always use t statistic for small or large samples):

$$t = \frac{\bar{x} - \mu_0}{\mathrm{SE}(\bar{x})} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

4. From the table for t distribution (Appendix C) with $(n - 1)$ degrees of freedom and the choice of α (e.g., $\alpha = 0.05$), the rejection region is determined by:
   a) For a one-sided test, use the column corresponding to an upper tail area of 0.05:

      t $\leq -$tabulated value for left tailed test and t $\geq$ tabulated value for right tailed test

   b) For a two-sided test or H_a: $\mu \neq \mu_0$, use the column corresponding to an upper tail area of $\alpha = 0.025$:

      t $\leq -$tabulated value for H_a: $\mu < \mu_0$ or t $\geq$ tabulated value H_a: $\mu > \mu_0$

This test is referred to as the one-sample t test.

Example 7.1

- Boys of a certain age have a mean weight of 85 lb.
- An observation was made that in a city neighborhood, children were underfed.
- As evidence, all 25 boys in the neighborhood of that age were weighed and found to have a mean $\bar{x}$ of 80.94 lb and a standard deviation s of 11.60 lb.
- An application of the procedure above yields

$$\mathrm{SE}(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{11.60}{\sqrt{25}} = 2.32$$

$$t = \frac{80.94 - 85}{2.32} = -1.75$$

The underfeeding complaint corresponds to the one-sided alternative

$$H_a\colon \mu < 85$$

so that we would reject the null hypothesis if t $\leq -$tabulated value.

From Appendix C and with 24 degrees of freedom $(n-1)$, we find that tabulated value $= 1.71$, under the column corresponding to a 0.05 upper tail area; the null hypothesis is rejected at the 0.05 level. In other words, there is enough evidence to support the underfeeding complaint.

Analysis of Pair-Matched Data (Two Dependent Samples)

- Applies to cases where each subject or member of a group is observed twice (e.g., before and after certain interventions) or matched pairs are measured for the same continuous characteristic.
- Data from matched (or before-and-after) experiments should never be considered as coming from two independent samples.
  - In this case, we reduce the data to a one-sample problem by computing before-and-after (or case-and control) difference for each subject or pairs of matched subjects.
  - We compare the means, before versus after or cases versus controls, and use of the sample of differences, $\{d_i\}$, one for each subject.
- The test statistic is the one-sample t test, the same one-sample t test as in the previous section:

$$t = \frac{\bar{d}-\mu_0}{S_d/\sqrt{n}},$$

where $\bar{d}$ and S_d are the sample mean differences and standard deviation respectively.

- The rejection region is determined using the t distribution at $n-1$ degrees of freedom.

Example 7.2

- Trace metals in drinking water affect the flavor of the water, and unusually high concentrations can pose a health hazard.
- Table 7.1 shows trace-metal concentrations (zinc, in mg/L) for both surface water and bottom water at six different river locations (the difference is bottom−surface).

- The necessary summarized figures are:

$$\bar{d} = \text{average difference}$$
$$= \frac{0.550}{6} = 0.0917 \text{ mg/L}$$
$$s_d^2 = \frac{0.068832 - (0.550)^2/6}{5} = 0.00368$$
$$s_d = 0.061$$
$$\text{SE}(\bar{d}) = \frac{0.061}{\sqrt{6}} = 0.0249$$
$$t = \frac{0.0917}{0.0249} = 3.68$$

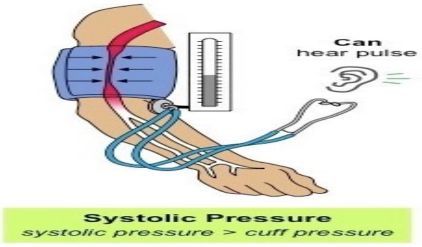
TABLE 7.1

Location	Bottom	Surface	Difference, d_i	d_i^2
1	0.430	0.415	0.015	0.000225
2	0.266	0.238	0.028	0.000784
3	0.567	0.390	0.177	0.030276
4	0.531	0.410	0.121	0.014641
5	0.707	0.605	0.102	0.010404
6	0.716	0.609	0.107	0.011449
Total			0.550	0.068832

- Using the column corresponding to the upper tail area of 0.025 in Appendix C, we have a tabulated value of 2.571 for 5 df.
- Since $t = 3.68 > 2.571$ we conclude that the null hypothesis of no difference should be rejected at the 0.05 level; *there is enough evidence to support the hypothesis of different mean zinc concentrations (two-sided alternative).*

Example 7.3

- The systolic blood pressures of $n = 12$ women between the ages of 20 and 35 were measured before and after administration of a newly developed oral contraceptive.
- Data are shown in Table 7.2 (the difference is after–before).
- The necessary summarized figures are



$$\bar{d} = \text{average difference} = \frac{31}{12} = 2.58 \text{ mmHg}$$

$$s_d^2 = \frac{185 - (31)^2/12}{11} = 9.54, \qquad s_d = 3.09$$

$$\mathrm{SE}(\bar{d}) = \frac{3.09}{\sqrt{12}} = 0.89$$

$$t = \frac{2.58}{0.89} = 2.90$$

TABLE 7.2

Subject	Systolic Blood Pressure (mmHg) Before	After	Difference, d_i	d_i^2
1	122	127	5	25
2	126	128	2	4
3	132	140	8	64
4	120	119	−1	1
5	142	145	3	9
6	130	130	0	0
7	142	148	6	36
8	137	135	−2	4
9	128	129	1	1
10	132	137	5	25
11	128	128	0	0
12	129	133	4	16
			31	185

- Using the column corresponding to the upper tail area of 0.05 in Appendix C, we have a tabulated value of 1.796 for 11 df.
- Since $t = 2.90 > 2.201$ we conclude that the null hypothesis of no blood pressure change should be rejected at the 0.05 level; there is enough evidence to support the hypothesis of increased systolic blood pressure (one-sided alternative).

Comparison of Two Mean (Independent Populations)

- When the population is normal and σ is unknown, where the sample size n is small, we will can use t-test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

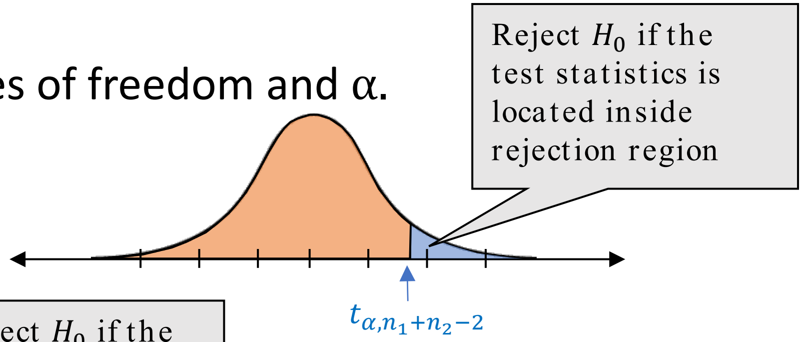
where $S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ is the pooled variance .

- The tabular value will be found from (Appendix C) with $(n_1 + n_2 - 2)$ degrees of freedom and α.

- One-sided (right tailed) test

$$H_0: \mu_1 = \mu_2 \text{ versus } H_a: \mu_1 > \mu_2$$

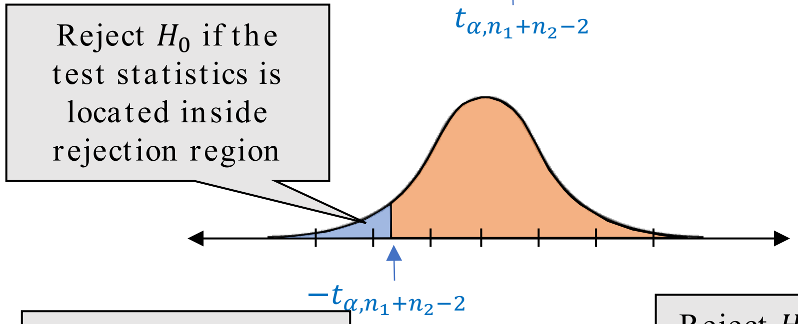
Reject H_0 *at the significance of level* α *if the* $t \geq t_{\alpha, n_1+n_2-2}$



- One-sided (left tailed) test

$$H_0: \mu_1 = \mu_2 \text{ versus } H_a: \mu_1 < \mu_2$$

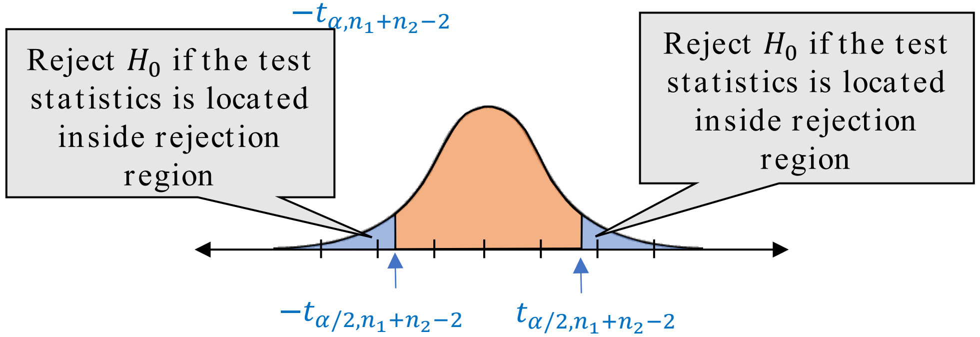
Reject H_0 *at the significance of level* α *if the* $t \leq -t_{\alpha, n_1+n_2-2}$



- Two-sided (two tailed) test

$$H_0: \mu_1 = \mu_2 \text{ versus } H_a: \mu_1 \neq \mu_2$$

Reject H_0 *at the level* α *if the* $t \geq t_{\alpha/2, n_1+n_2-2}$ *or* $t \leq t_{\alpha/2, n_1+n_2-2}$



Example 7.5

- In an attempt to assess the physical condition of joggers, a sample of $n_1 = 25$ joggers was selected and their maximum volume of oxygen uptake (VO2) was measured with the following results:

$$\bar{x}_1 = 47.5 \text{ mL/kg} \qquad s_1 = 4.8 \text{ mL/kg}$$

Results for a sample of $n_2 = 26$ nonjoggers were

$$\bar{x}_2 = 37.5 \text{ mL/kg} \qquad s_2 = 5.1 \text{ mL/kg}$$

To proceed with the two-tailed, two-sample t test, we have

$$\begin{aligned} s_p^2 &= \frac{(24)(4.8)^2 + (25)(5.1)^2}{49} \\ &= 24.56 \\ s_p &= 4.96 \end{aligned}$$

$$\begin{aligned} \mathrm{SE}(\bar{x}_1 - \bar{x}_2) &= 4.96\sqrt{\frac{1}{25} + \frac{1}{26}} \\ &= 1.39 \end{aligned}$$

It follows that

$$\begin{aligned} t &= \frac{47.5 - 37.5}{1.39} \\ &= 7.19 \end{aligned}$$

indicating a significant difference between joggers and nonjoggers (at 49 degrees of freedom and $\alpha = 0.01$, the tabulated t value, with an upper tail area of 0.025, is about 2.0).

Example 7.6

- Vision, or more especially visual acuity, depends on a number of factors.
- A study was undertaken in Australia to determine the effect of one of these factors: racial variation.
- Visual acuity of recognition as assessed in clinical practice has a defined normal value of 20/20 (or zero in log scale).
- The following summarized data on monocular visual acuity (expressed in log scale) were obtained from two groups:

- **Australian males of European origin**

$$n_1 = 89 \quad \bar{x}_1 = -0.20 \quad s_1 = 0.18$$

- **Australian males of Aboriginal origin**

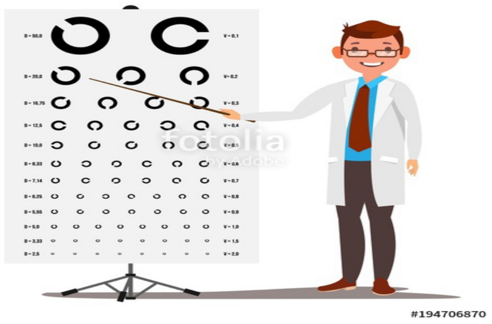
$$n_2 = 107 \quad \bar{x}_2 = -0.26 \quad s_2 = 0.13$$

To proceed with a two-sample t test, we have

$$s_p^2 = \frac{(88)(0.18)^2 + (106)(0.13)^2}{194} = (0.155)^2$$

$$\mathrm{SE}(\bar{x}_1 - \bar{x}_2) = (0.155)\sqrt{\frac{1}{89} + \frac{1}{107}} = 0.022$$

$$t = \frac{(-0.20) - (-0.26)}{0.022} = 2.73$$



The result indicates that the difference is statistically significant beyond the 0.01 level (at $\alpha = 0.01$, and for a two-sided test the cut point is 2.58 for high df values).

Example 7.7

- The extend to which an infant's health is affected by parental smoking is an important public health concern.
- The following data are the urinary concentrations of cotinine (a metabolite of nicotine);
- measurements were taken both from a sample of infants who had been exposed to household smoke and from a sample of unexposed infants.
- The statistics needed for our two-sample t test are:

Unexposed ($n_1 = 7$)	8	11	12	14	20	43	111	
Exposed ($n_2 = 8$)	35	56	83	92	128	150	176	208

- **For unexposed infants:**

$$n_1 = 7 \quad \bar{x}_1 = 31.29 \quad s_1 = 37.07$$

- **For exposed infants:**

$$n_2 = 8 \quad \bar{x}_2 = 116.00 \quad s_2 = 59.99$$

To proceed with a two-sample t test, we have

$$s_p^2 = \frac{(6)(37.07)^2 + (7)(59.99)^2}{13} = (50.72)^2$$

$$\mathrm{SE}(\bar{x}_1 - \bar{x}_2) = (50.72)\sqrt{\frac{1}{7} + \frac{1}{8}} = 26.25$$

$$t = \frac{116.00 - 31.29}{26.25} = 3.23$$

The result indicates that the difference is statistically significant beyond the 0.01 level (at $\alpha = 0.01$, and for a two-sided test the cut point is 3.012 for 13 df).

Nonparametric Methods

- **Parametric versus nonparametric methods**
- Recall that the **parametric tests** (z score or t-test) depend on certain assumptions about distributions in the population. Departures from these assumptions will biased the results.
- On the other hand, the **nonparametric tests** are robust where departures from these assumptions have very little effect on the results, provided that samples are large enough.
- Both methods (Parametric and nonparametric) sensitive to extreme observations, a few very small or very large—perhaps erroneous—data values.

	Parametric test	Nonparametric test
One sample	One-Sample t –Test	Signed-Rank Test
Pair-matched samples	One-Sample t –Test	Wilcoxon Signed-Rank Test
Independent samples	Two-Samples t –Test	Wilcoxon Rank-Sum Test

Wilcoxon Rank-Sum Test (Mann Whitney U Test)

- **Wilcoxon test** (also known as **Mann Whitney U Test**) is a nonparametric counterpart of the two-sample t test.
- Used to compare two samples that have been drawn from independent populations.
- unlike the t test, the Wilcoxon test does not assume that the underlying populations are normally distributed and is less affected by extreme observations.
- The Wilcoxon rank-sum test evaluates the null hypothesis that the medians of the two populations are identical (H_0: two medians are equal vs H_a: two medians are not equal).
    - For a normally distributed population, the population median is also the population mean.

Wilcoxon Rank-Sum Test Steps

- Consider the following example:
- A study was designed to test the question of whether cigarette smoking is associated with reduced serum-testosterone levels.
- Two samples, each of size 10, are selected independently.
- The first sample consists of 10 nonsmokers who have never smoked.
- The second sample consists of 10 heavy smokers, defined as those who smoke 30 or more cigarettes a day.
- To perform the Wilcoxon rank-sum test

1. We combine the two samples into one large sample (of size 20),
2. arrange the observations from smallest to largest, and assign a rank, from 1 to 20, to each.
    - If there are tied observations, we assign an average rank to all measurements with the same value. (For example, if the two observations next to the third smallest are equal, we assign an average rank of $(4+5)/2=4.5$ to each one.
3. Find the sum of the ranks corresponding to each of the original samples.

Wilcoxon Rank-Sum Test

- To test the null hypothesis that the two underlying populations have identical medians, we use the statistic

$$Z = \frac{R - \mu_R}{\sigma_R},$$

  - $\mu_R = \frac{n_1(n_1+n_2+1)}{2}$ is the mean of R,
  - $\sigma_R = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}$ is the standard deviation of R,
  - n_1 and n_2 are the two sample sizes,
  - R is the sum of the ranks from the sample with size n_1.
- For relatively large values of n_1 and n_2 (say ≥ 10) the sampling distribution of this statistic is approximately standard normal.
- The null hypothesis is rejected at the 5% level, against a two-sided alternative, if

$$\mathrm{Z} \geq 1.96 \text{ or } Z \leq -1.96.$$

Example 7.8

- For the study on cigarette smoking, Table 7.4 shows the raw data,
- Testosterone levels were measured in mg/dL and the ranks were determined.

- The sum of the ranks for group 1 (nonsmokers) is

$$R = 143$$

- In addition,

$$\mu_R = 10(10 + 10 + 1)/2 = 105,$$

$$\sigma_R = \sqrt{10 \times 10(10 + 10 + 1)/12} = 13.23$$

- Substituting these values into the equation for the test statistic, we have

$$Z = \frac{R - \mu_R}{\sigma_R} = \frac{143 - 105}{13.23} = 2.87$$

TABLE 7.4

Nonsmokers		Heavy Smokers	
Measurement	Rank	Measurement	Rank
0.44	8.5	0.45	10
0.44	8.5	0.25	1
0.43	7	0.40	6
0.56	14	0.27	2
0.85	17	0.34	4
0.68	15	0.62	13
0.96	20	0.47	11
0.72	16	0.30	3
0.92	19	0.35	5
0.87	18	0.54	12

- Since z > 1.96, we reject the null hypothesis at the 5% level. (In fact, since z > 2.58, we reject the null hypothesis at the 1% level.)
- Note that if we use the sum of the ranks for the other group (heavy smokers), the sum of the ranks is 67, leading to a z score of $Z = (67 - 105)/13.23 = -2.87$, and we would come to the same decision.

Example 7.9

- Refer to the nicotine data of Example 7.7, where measurements were taken both from a sample of infants who had been exposed to household smoke and from a sample of unexposed infants. We have:

Unexposed ($n_1 = 7$)	8	11	12	14	20	43	111
Rank	1	2	3	4	5	7	11

Exposed ($n_2 = 8$)	35	56	83	92	128	150	176	208
Rank	6	8	9	10	12	13	14	15

The sum of the ranks for the group of exposed infants is

$$R = 87$$

In addition,

$$\mu_R = \frac{(8)(8+7+1)}{2} = 64$$

and

$$\sigma_R = \sqrt{\frac{(8)(7)(8+7+1)}{12}} = 8.64$$

Substituting these values into the equation for the Wilcoxon test, we have

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{87 - 64}{8.64} = 2.66$$

Since $z > 1.96$, we reject the null hypothesis at the 5% level. In fact, since $z > 2.58$, we reject the null hypothesis at the 1% level; p value < 0.01. (It should be noted that the sample sizes of 7 and 8 in this example may be not large enough.)

Wilcoxon Signed-Rank Test

- The idea of using ranks, instead of measured values, to form statistical tests to compare population means applies to the analysis of pair-matched data as well.
- As with the one-sample t test for pair-matched data, we begin by forming differences.
- Then the absolute values of the differences are assigned ranks; if there are ties in the differences, the average of the appropriate ranks is assigned.
- Next, we attach a $+$ or a $-$ sign back to each rank, depending on whether the corresponding difference is positive or negative.
  - This is achieved by multiplying each rank by $+1$, -1, or 0 as the corresponding difference is positive, negative, or zero.
  - The results are n signed ranks, one for each pair of observations; for example, if the difference is zero, its signed rank is zero.
- The basic idea is that if the mean difference is positive, there would be more and larger positive signed ranks; since if this were the case, most differences would be positive and larger in magnitude than the few negative differences, most of the ranks, especially the larger ones, would then be positively signed.

Wilcoxon Signed-Rank Test

- To test the null hypothesis that the two underlying populations have identical means, we use the statistic

$$Z = \frac{R - \mu_R}{\sigma_R},$$

  - $\mu_R = \frac{n(n+1)}{4}$ is the mean of R,
  - $\sigma_R = \sqrt{\frac{n(n+1)\,(2n+1)}{24}}$ is the standard deviation of R,
  - R is the sum of the positive signed ranks.
- For relatively large values of n ($n \geq 20$) the sampling distribution of this statistic is approximately standard normal.
- The null hypothesis is rejected at the 5% level, against a two-sided alternative, if

$$\mathrm{Z} \geq 1.96 \text{ or } Z \leq -1.96.$$

Example 7.10

- Ultrasounds were taken at the time of liver transplant and again 5 to 10 years later to determine the systolic pressure of the hepatic artery.
- Results for 21 transplants for 21 children are shown in Table 7.5.
- The sum of the positive signed ranks is

$$13 + 9 + 15.5 + 20 + 5.5 + 4 + 5.5 + 17.5 = 90$$

Its mean and standard deviation under the null hypothesis are

$$\mu_R = \frac{21 \times 22}{4} = 115.5 \text{ and } \sigma_R = \sqrt{\frac{21(22)\,(43)}{24}} = 28.77$$

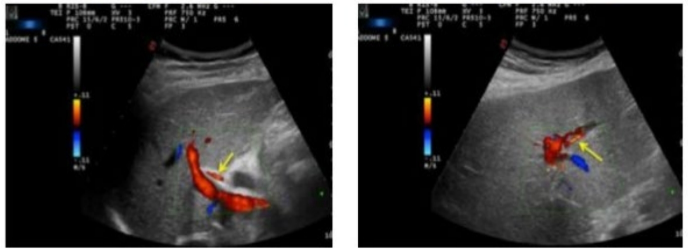
leading to a standardized z score of

$$Z = \frac{90 - 115.5}{28.77} = -0.89$$

The result indicates that the systolic pressure of the hepatic artery measured five years after the liver transplant, compared to the measurement at transplant, is lower on the average; however, the difference is not statistically significant at the 5% level (-0.89 > -1.96).

Normal hepatic artery

Right hepatic artery | Left hepatic artery



Intercostal approach | Epigastric approach

TABLE 7.5

Child	Later	At Transplant	Difference	Absolute Value of Difference	Rank	Signed Rank
1	46	35	11	11	13	13
2	40	40	0	0	2	0
3	50	58	−8	8	9	−9
4	50	71	−19	19	17.5	−17.5
5	41	33	8	8	9	9
6	70	79	−9	9	11	−11
7	35	20	15	15	15.5	15.5
8	40	19	21	21	20	20
9	56	56	0	0	2	0
10	30	26	4	4	5.5	5.5
11	30	44	−14	14	14	−14
12	60	90	−30	30	21	−21
13	43	43	0	0	2	0
14	45	42	3	3	4	4
15	40	55	−15	15	15.5	−15.5
16	50	60	−10	10	12	−12
17	66	62	4	4	5.5	5.5
18	45	26	19	19	17.5	17.5
19	40	60	−20	20	19	−19
20	35	27	−8	8	9	−9
21	25	31	−6	6	7	−7

One-Way Analysis of Variance (ANOVA)

- The one-way analysis of variance (ANOVA) is used to test the null hypothesis that three or more population means are equal (*simultaneously compare these means in one step*) against the alternative hypothesis that at least two means are differ.
- We have continuous measurements X's from k independent samples;
- The sample sizes may or may not be equal.
- We assume that these are samples from k normal distributions with a common variance σ^2, but the means, μ_i's, may or may not be the same.
- ANOVA is a method simultaneously test the null hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ against the alternative hypothesis $H_a: \mu_i \neq \mu_j$ for at least one pair $i \neq j$.
- ANOVA is an extension of the two independent samples t-test (k=2).

One-Way Analysis of Variance (ANOVA)

Treatment 1	Treatment 2	· · ·	Treatment k
x_{11}	x_{21}		x_{k1}
x_{12}	x_{22}		x_{k2}
$\vdots$	$\vdots$		$\vdots$
x_{1n_1}	x_{2n_2}		x_{kn_k}
$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1}$	$\bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2}$		$\bar{x}_k = \frac{\sum_{i=1}^{n_k} x_{ki}}{n_k}$
$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n}$, where $n = n_1 + n_2 + \cdots + n_k$			

- Data from the ith sample can be summarized into sample size n_i, sample mean $\bar{x}_i$, and sample variance S_i^2, where x_{ij} is the jth measurement from the ith sample.
- Grand Mean = $\bar{x}$ is the average of all observations (pooled data). It is a weighted average of the individual sample means.

One-Way ANOVA and the Variation in $\boldsymbol{X}$

- The total variation, denoted by **SST**, is the sum of squared deviations:

$$\text{SST} = \sum_{i,j} (x_{ij} - \bar{x})^2$$

- The total variation in the combined sample can be decomposed into two components

$$x_{ij} - \bar{x} = (x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{x})$$

1. The first term reflects the variation within the ith sample; the sum

$$\text{SSW} = \sum_{i,j} (x_{ij} - \bar{x}_i)^2 = \sum_i (n_i - 1) s_i^2$$

   is called the *within sum of squares*.

2. The difference between the two sums of squares above,

$$\text{SSB} = \text{SST} - \text{SSW} = \sum_{i,j} (\bar{x}_i - \bar{x})^2 = \sum_i n_i (\bar{x}_i - \bar{x})^2$$

   is called the *between sum of squares*.

SSB represents the variation or differences between the sample means, a measure very similar to the numerator of a sample variance; the n_{i} values serve as weights.

One-Way ANOVA Table

- To test $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$, the F-test statistic (Table 7.6) is used and H_0 rejected at level α if F-statistic $\geq$ tabular F value found in table (Appendix E) with $(k-1, n-k)$ degrees of freedom.

1. The *within mean square*

$$\text{MSW} = \frac{\text{SSW}}{n-k} = \frac{\sum_i (n_i - 1)s_i^2}{\sum (n_i - 1)}$$

serves as an estimate of the common variance σ^2 as stipulated by the one-way ANOVA *model*. In fact, it can be seen that MSW is a natural extension of the pooled estimate s_p^2 as used in the two-sample t test; It is a measure of the average variation within the k samples.

2. The *between mean square*

$$\text{MSB} = \frac{\text{SSB}}{k-1}$$

represents the *average* variation (or differences) between the k sample means.

TABLE 7.6

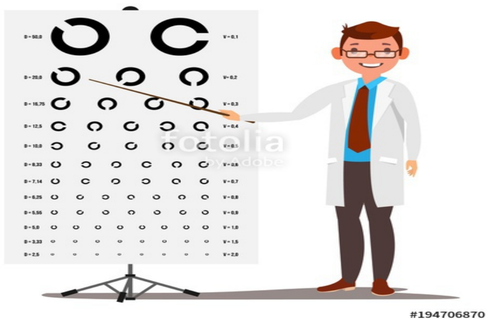
Source of Variation	SS	df	MS	F Statistic	p Value
Between samples	SSB	$k-1$	MSB	MSB/MSW	p
Within samples	SSW	$n-k$	MSW		
Total	SST	$n-1$			

In fact, when $k = 2$, we have

$$F = t^2$$

where t is the test statistic for comparing the two population means. In other words, when $k = 2$, the F test is equivalent to the two-sided two-sample t test.

Example 7.11



- Vision, especially visual acuity, depends on a number of factors (See Example 7.6).
- A study was undertaken in Australia to determine the effect of one of these factors: racial variation.
- Visual acuity of recognition as assessed in clinical practice has a defined normal value of 20=20 (or zero on the log scale).
- The following summarize the data on monocular visual acuity (expressed on a log scale).

1. Australian males of European origin

$$n_1 = 89$$
$$\bar{x}_1 = -0.20$$
$$s_1 = 0.18$$

2. Australian males of Aboriginal origin

$$n_2 = 107$$
$$\bar{x}_2 = -0.26$$
$$s_2 = 0.13$$

3. Australian females of European origin

$$n_3 = 63$$
$$\bar{x}_3 = -0.13$$
$$s_3 = 0.17$$

4. Australian females of Aboriginal origin

$$n_4 = 54$$
$$\bar{x}_4 = -0.24$$
$$s_4 = 0.18$$

Example 7.11 (Cont.)

- To proceed with a one-way analysis of variance, we calculate the mean of the combined sample

$$\bar{x} = \frac{(89)(-0.20) + (107)(-0.26) + (63)(-0.13) + (54)(-0.24)}{89 + 107 + 63 + 54} = -0.213$$

$$\begin{aligned} \text{SSB} &= (89)(-0.20 + 0.213)^2 + (107)(-0.26 + 0.213)^2 \\ &\quad + (63)(-0.13 + 0.213)^2 + (54)(-0.24 + 0.213)^2 \\ &= 0.7248 \\ \text{MSB} &= \frac{0.7248}{3} \\ &= 0.2416 \\ \text{SSW} &= (88)(0.18)^2 + (106)(0.13)^2 + (62)(0.17)^2 + (53)(0.18)^2 \\ &= 8.1516 \\ \text{MSW} &= \frac{8.1516}{309} \\ &= 0.0264 \\ F &= \frac{0.2416}{0.0264} \\ &= 9.152 \end{aligned}$$

- The results are summarized in an ANOVA table (Table 7.7).
- The resulting F test indicates that the overall differences between the four population means is highly significant ($p < 0.00001$).

TABLE 7.7

Source of Variation	SS	df	MS	F Statistic	p Value
Between samples	0.7248	3	0.2416	9.152	<0.0001
Within samples	8.1516	309	0.0264		
Total	8.8764	312			

Example 7.12

- A study was conducted to test the question as to whether cigarette smoking is associated with reduced serum-testosterone levels in men aged 35 to 45.
- The study involved the following four groups:
  1. Nonsmokers who had never smoked,
  2. Former smokers who had quit for at least six months prior to the study,
  3. Light smokers, defined as those who smoked 10 or fewer cigarettes per Day,
  4. Heavy smokers, defined as those who smoked 30 or more cigarettes per day.
- Each group consisted of 10 men and Table 7.8 shows raw data, where serum-testosterone levels were measured in mg/dL.

TABLE 7.8

Nonsmokers	Former Smokers	Light Smokers	Heavy Smokers
0.44	0.46	0.37	0.44
0.44	0.50	0.42	0.25
0.43	0.51	0.43	0.40
0.56	0.58	0.48	0.27
0.85	0.85	0.76	0.34
0.68	0.72	0.60	0.62
0.96	0.93	0.82	0.47
0.72	0.86	0.72	0.70
0.92	0.76	0.60	0.60
0.87	0.65	0.51	0.54

- An application of the one-way ANOVA yields Table 7.9.
- The resulting F test indicates that the overall differences between the four population means is statistically significant at the 5% level but not at the 1% level ($p = 0.0179$).

TABLE 7.9

Source of Variation	SS	df	MS	*F* Statistic	*p* Value
Between samples	0.3406	3	0.1135	3.82	0.0179
Within samples	1.0703	36	0.0297		
Total	1.4109	39			