Tumor number is measured in 18 semi-resistant rats and in 13 susceptible rats.

Results: $n_1 = 18$, $\bar{x} = 7.53$, $s_1^2 = 15.876$
$n_2 = 13$, $\bar{y} = 11.767$, $s_2^2 = 17.161$.

Using a two-sided alternative, test the null hypothesis that the mean tumor number in the two populations of rats are the same ($\alpha = 0.05$). Assume the tumor numbers are normally distributed in both populations, and assume the the population variances are equal.

$H_0 : \mu_1 = \mu_2$ vs $H_A : \mu_1 \neq \mu_2$.

Evaluate the test statistic:

$$= \frac{7.53 - 11.767}{\sqrt{16.408/18 + 16.408/13}} = -2.874.$$

The critical value is $t_{29,0.025}$, which is 2.045. Thus, we reject the null at a significance level of 0.05. Note that the p-value is
$P(T_{29} \leq -2.874) + P(T_{29} \geq 2.874) = 0.0037 + 0.0037 = 0.0074$

What if we didn't assume (for the last calculation) that the variances were equal ?

Was the normality assumption for the last example a good one ?

NONPARAMETRIC STATISTICS

We will consider methods for statistical inference which do not depend upon knowledge of the functional form of the underlying probability distributions. The primary advantage of the approaches we will discuss is that they are **distribution-free**, in that they make no assumptions about the sample population.

A powerful distribution-free tool is the use of **ranks**. The rank of an observation is the relative position of an observation's magnitude compared to the rest of the sample. When two or more observations have the same value (ties), the rank is assigned by computing the average of the ranks that would have been assigned to tied values and using this average as the common rank shared by each of the tied values.

As an example, suppose my sample consisted of the 5 observations $\{-4.3, 2.7, 0.6, -5.6, 7\}$, the respective ranks would be $\{2, 4, 3, 1, 5\}$. The easiest way to do this is to order the values first and then assign ranks.

For another example, suppose my sample consists of the 9 observations $\{1.3, 2.7, -3.2, 0.7, 0.7, -5.6, 3.1, 3.1, 3.1\}$. The ordered observations and ranks are as follows:

positions	1	2	3	4	5	6	7	8	9
observations	-5.6	-3.2	0.7	0.7	1.3	2.7	3.1	3.1	3.1
rank	1	2	3.5	3.5	5	6	8	8	8

Amazingly, if we consider only continuous distributions (to avoid ties), the distribution of ranks *does not depend on the particular continuous distribution of the sample*. In other works, rank based procedures are distribution-free.

The Wilcoxon signed rank test

Sometimes we wish to test a null hypothesis about a population mean, but if the sample size is small and we have nonnormally distributed variables, the t-test may not be appropriate. Here we discuss a rank based procedure for testing the null hypothesis that the median difference between 2 random variables is 0.

The *Wilcoxon signed rank test* is appropriate for paired data and other kinds of one-sample data consisting of independent observations and where the null hypothesis is that in the underlying population of differences, the median difference is zero.

To conduct the Wilcoxon signed rank test:

- Take the differences between all paired observations.
- Rank the absolute differences from smallest to largest.
- Add up the ranks associated with positive differences and add up the ranks associated with negative differences to obtain T_+ and T_-, respectively. Note that ranks of zero difference are not used. In this case, the sample size is reduced by the number of differences that are zero.
- Let T denote the smaller sum.
- T is the test statistic. Table A.6 gives $P(T \leq t)$ for different values of n and t.
- Using Table A.6, calculate the p-value.

Consider a study which taught people how to brush their teeth. Twelve pairs of patients were obtained by carefully matching on factors such as age, sex, intelligence, and initial oral hygiene scores. One member of each pair received instruction on how to brush their teeth and on other oral hygiene matters. Six months later all 24 subjects were examined and assigned an oral hygiene score by a dental hygienist unaware of which subjects had received the instruction. A low score indicated a high level of dental hygiene.

pair number	Instructed	not instructed	sign of difference	rank
1	1.50	2.0	-	5.5
2	2.00	2.30	-	2.5
3	3.60	4.00	-	4
4	3.00	2.50	+	5.5
5	3.50	2.20	+	9
6	2.50	3.10	-	7
7	2.30	3.80	-	11.5
8	4.00	2.60	+	10
9	1.60	3.10	-	11.5
10	2.90	3.20	-	2.5
11	3.30	3.20	+	1
12	2.20	2.90	-	8

The total of the ranks associated with negative values is

$$t_- = 5.5 + 2.5 + 4 + 7 + 11.5 + 11.5 + 2.5 + 8 = 52.5$$

The total of the ranks associated with positive values is

$$t_+ = 5.5 + 9 + 10 + 1 = 25.5$$

The smaller of these, $t_+ = 25.5$, is then the value of the Wilcoxon signed rank test statistic. In Table A.6, we find that for a sample of size 12, $P(T \leq 25) = 0.1507$. Thus, the p-value is 2*0.1507=0.3014. We do not reject H_0 at $\alpha = 0.05$.

Another example: In a placebo (P) controlled clinical trial of a revolutionary drug (D) for treating Alzheimer's disease, 12 Alzheimer's patients were assigned to one of two regimens:

A: P for two months followed by one month with no medication followed by D for two months or

B: D for two months followed by one month with no medication followed by P for two months.

A mental performance measure (ranging from 0–30, the higher the better) was assessed at 2 months and at 5 months, reflecting the effect of either P or D, depending on the regimen. Test $H_0 : \mu_d = 0$ versus a two-sided alternative at significance level $\alpha = 0.05$.

The resulting scores and signed ranks are as follows:

	patient										
drug	1	2	3	4	5	6	7	8	9	10	11
P	12	9	7	24	28	13	8	11	9	23	10
D	16	11	13	25	25	21	7	16	18	30	9
P-D	-4	-2	-6	-1	3	-8	1	-5	-9	-7	1
|P-D|	4	2	6	1	3	8	1	5	9	7	1
ranks	6	4	8	2	5	10	2	7	11	9	2

The total of the ranks associated with negative values is

$$t_- = 6 + 4 + 8 + 2 + 10 + 7 + 11 + 9 = 57.$$

The total of the ranks associated with positive values is

$$t_+ = 5 + 2 + 2 = 9$$

The smaller of these, $t_+ = 9$, is then the value of the Wilcoxon signed rank test statistic.

In Table A.6, we find that for a sample of size 12, $P(T \leq 9) = 0.0081$. Thus, the p-value is 2*0.0081=0.0162. We would reject H_0 at $\alpha = 0.05$ and conclude that the treatment is effective.

The Wilcoxon Rank Sum test

The Wilcoxon Rank Sum test is used for comparing the medians of two samples. The null hypothesis is $H_0 : M_x = M_y$, and either one or two-sided alternatives may be used.

The Wilcoxon Rank Sum test (Mann - Whitney U test):

1. Used to compare two samples that have been drawn from independent populations.
2. It is a non-parametric counterpart to the two sample t-test.
3. It doesn't require that the underlying populations be normally distributed or that their variances be equal.

To conduct the Wilcoxon Rank Sum test:

- Let n_1 be the sample size of sample 1 (X) and let n_2 be the sample size of sample 2 (Y).
- Combine the $n_1 + n_2 = n$ data values together and compute the ranks of the combined sample.
- Add up the ranks associated with sample 1 and denote this W_1; do the same for sample 2 and denote this W_2.
- Let W denote the smaller sum.
- W is the test statistic. Table A.7 gives $P(W \leq w)$ for different values of n_1, n_2, and w.
- Using Table A.7, calculate the p-value.

Suppose we wish to compare the length of stay in the hospital for patients with the same diagnosis at two different hospitals.

Hospital A: 21, 10, 32, 60, 8, 44, 29,

Hospital B: 86, 27, 10, 68, 87, 76, 125, 60, 96

Test the null hypothesis that the median length of stay at the two hospitals is identical, against a two-sided alternative hypothesis at level $\alpha = 0.05$.

Ranks obtained from combined sample

Hospital A: 4, 2.5, 7.0, 9.5, 1.0, 8.0, 6.0

Hospital B: 12.0, 5.0, 2.5, 11.0, 14.0, 12.0, 16.0, 9.5, 15.0

In this problem, $n_1 = 7$, $n_2 = 9$, and the sum of the ranks from hospitals A and B are 38 and 98, respectively.

The value of the test statistic is $w = 38$.
From Table A.7 we see that $P(W \leq 38) = 0.0115$ and so the p-value for this two-sided test is 0.0230. We reject the null hypothesis for significance level $\alpha = 0.05$.

Note

The Wilcoxon signed rank test is a non-parametric version of the paired t-test. The Wilcoxon rank sum test is a non-parametric version of the two (independent) sample t-test.

These non-parametric tests are used when the sample size is very small or when the normality assumption is violated.

Inference on more than two means

Suppose we are interested in measuring FEV and we collect FEV recordings form three groups of patients (21 patients from Johns Hopkins, 16 patients from Rancho Los Amigos (RLA), and 23 patients from St. Louis). We are interested in testing

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

against the alternative that at least one of the population means differs from one of the others.

Results:

Group i	n_i	$\bar{x}_i$	s_i
1	21	2.63	0.496
2	16	3.03	0.523
3	23	2.88	0.498

Idea: Do 3 t-tests comparing group 1 to group 2, group 2 to group 3, and group 1 to group 3.

Problem with that idea: Type I error probability gets large.

For example, if for each of the three t-tests, we set $\alpha = 0.05$, then the type I error probability for all 3 tests is the probability of rejecting at least one out of the three tests, given that the null is true.

$$\begin{aligned} & P(\text{rejecting at least one} \mid \text{null is true}) \\ & \quad = 1 - P(\text{rejecting none of the three} \mid \text{null is true}) \\ & \quad = 1 - (0.95^3) \\ & \quad = 1 - 0.857 \\ & \quad = 0.143 \end{aligned}$$