Nonparametric Two-Sample Tests

Sign test

Mann-Whitney U-test

(a.k.a. Wilcoxon two-sample test)

Kolmogorov-Smirnov Test

Wilcoxon Signed-Rank Test

Tukey-Duckworth Test

Nonparametric Tests

Recall, nonparametric tests are considered "distribution-free" methods because they do not rely on any underlying mathematical distribution.

They do, however, have various assumptions that must be met.

Do not be confused by not having the need to meet an assumption of "normality" with the notion of "assumptionless."

Sign Test

Consider a simple example where 20 farmers are given two fertilizers (A & B) by an extension agent and asked to evaluate which one was "better".

In the end, 16 farmers reported that fertilizer A was the better of the two (based on qualitative estimates of visual quality, greenness, yield, etc.).

If there were no difference between fertilizers, we would expect fertilizer A to be binomially distributed with P = 0.05 and N = 20.

To estimate our rejection region:

```
> pbinom(16,size=20,prob=0.5)
[1] 0.9987116
```

Intuitively, we would expect to subtract 1.0 to get the other tail, but we must actually adjust for the fact that one tail is 16 or better and for counts, the other tail must be 15 or fewer:

```
> 1-pbinom(15,size=20,prob=0.5)
[1] 0.005908966
```

If you wish a 2-tailed test, you need to add the probabilities in each direction:

```
> 1-pbinom(15,20,0.5)+pbinom(4,20,0.5)
[1] 0.01181793
```


Sign Test

This is a bit of a confusing test to start out with (we do so only because it is one of the oldest statistical tests*) because we are assuming a binomial distribution.

If we are assuming a distribution, then are we not doing a parametric test? Yes. BUT, in this case the parametric binomial distribution and the C distribution of the sign test are identical. You are only ever working with two outcomes: A/B, dead/alive, 0/1, +/-, etc. The latter is where the “sign test” originated from.

* First discovered in late 1600s

Mann-Whitney *U*-Test

This is the nonparametric analog to the two-sample *t*-test with equal variances.

It is used primarily when the data have not met the assumption of normality (or should be used when there is sufficient doubt).

Assumptions:
- Independent samples
- Continuous variable
- Equal variances
- Identical (non-normal) distributions

Mann-Whitney *U*-Test

This test is based on ranks.

It has good efficiency, especially for symmetric distributions.

There are exact procedures for this test given small samples with no ties, and there are large sample approximations.

The Mann-Whitney test statistic, U, is defined as the total number of times a Y_1 precedes a Y_2 in the configuration of combined samples.

Mann-Whitney *U*-Test

- Procedure -

1. Pool data together, sort data in ascending order, keep track of sample ID
2. Convert data to ranks (1, 2, 3,... Y)
3. Separate ranks back in to two samples
4. Compute the test statistic, U
5. Determine critical value of U from table
6. Formulate decision and conclusion

Mann-Whitney *U*-Test

- Example -

Suppose you wished to determine if there was a difference in the biomass of male and female Juniper trees.

Thus, H_o: $B_{male} = B_{female}$ (medians are equal)
H_a: $B_{male} \neq B_{female}$ (medians not equal)

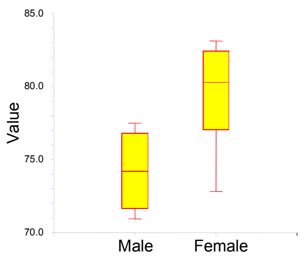
You randomly select 6 individuals of each gender from the field, dry them to constant moisture, chip them, and then weigh them to the nearest kg.

Mann-Whitney *U*-Test

- Example -

Raw Data:

Male	74	77	78	75	72	71
Fem	80	83	73	84	82	79



Preliminary analysis shows data to have equal variances, but normality tests are questionable given small sample sizes.

Mann-Whitney U-test is most appropriate...

Mann-Whitney *U*-Test

- Example -

Raw Data:

Male	$\underline{74}$	$\underline{77}$	$\underline{78}$	$\underline{75}$	$\underline{72}$	$\underline{71}$
Fem	80	83	73	84	82	79

Order & Rank, ID Sample by Underline:

$\underline{71}$ $\underline{72}$ 73 $\underline{74}$ $\underline{75}$ $\underline{77}$ $\underline{78}$ 79 80 82 83 84

$\underline{1}$ $\underline{2}$ 3 $\underline{4}$ $\underline{5}$ $\underline{6}$ $\underline{7}$ 8 9 10 11 12

Sum the Ranks:

Male: 1 + 2 + 4 + 5 + 6 + 7 = 25

Fem: 3 + 8 + 9 + 10 + 11 + 12 = 53

Mann-Whitney *U*-Test

- Example, cont. -

Compute test statistic using rank data.
First calculate *C*, then compare to ($n_1 n_2$ -*C*).
Larger of two values becomes U statistic.

$$C = n_1 n_2 + n_2 \frac{(n_2+1)}{2} - \sum_{i=1}^{n_2} R_i$$

n_1 = *N* of larger sample
n_2 = *N* of smaller sample
R = Ranks of smaller sample

In our example, $C = (6)(6) + (6)(7)/2 - 25 = 32$
$n_1 n_2 - C = (6)(6) - 32 = 4$
Larger value becomes $U_{calc} = 32$
$U_{6,6,\ .025} = U_{table} = 31$
$U_{calc} > U_{table}$, therefore reject H_o

Mann-Whitney *U*-Test

The only real hitch that can arise with this test is when two (or more) ranks are tied. When this happens, sum the ranks and give each the mean rank value.

Example: 24, 24 tied at 3, 4 then 3.5, 3.5

Note also that at N > 20, *U* begins to approximate *t*, so the test statistic changes to a *t*-value:

$$t_s = \frac{\left(U_s - \frac{n_1 n_2}{2}\right)}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Mann-Whitney *U*-Test
- Using R -

```
> male<-c(74,77,78,75,72,71)
> female<-c(80,83,73,84,82,79)

> wilcox.test(male,female) #NB: = MWU

        Wilcoxon rank sum test

data:  male and female
W = 4, p-value = 0.02597
alternative hypothesis: true location
shift is not equal to 0
```

What would have happened if one had "mis-applied" the *t*-test instead of using the Mann-Whitney *U*-test?

Both samples would pass a normality test (but would be questionable given the small sample size) and both would pass a homogeneity of variance test. The result will be the same, but note the difference in *P*-value.

```
> t.test(male,female,var.equal=TRUE)

       Two Sample t-test

data:  male and female
t = -2.8776, df = 10, p-value = 0.01645
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.054334  -1.278999
sample estimates:
mean of x mean of y
 74.50000  80.16667
```

Kolmogorov-Smirnov Test

This is the nonparametric analog to the two-sample t-test with unequal variances.

It is often used when the data have not met either the assumption of normality or the assumption of equal variances.

Assumptions:
- Variable at least ordinal
- Two samples are independent
- Both simple random samples
- Identical distributions

Kolmogorov-Smirnov Test

This test has poor statistical efficiency.

Many nonparm stats are based on ranks and therefore measure differences in location. The K-S examines a single maximum difference between two distributions.

If a statistical difference is found between the distributions of X and Y, the test provides no insight as to what caused the difference. The difference could be due to differences in location (mean), variation (standard deviation), presence of outliers, type of skewness, type of kurtosis, number of modes, and so on.

Kolmogorov-Smirnov Test
- Procedure -

Note that the hypotheses for K-S are NOT rooted in a mean or median (measures of central tendency).

The null and alternative hypotheses for the K-S test relate to the equality of the two distribution functions [usually noted as F(X) or F(Y)].

Thus, the typical two-tailed hypothesis becomes:

H_o: $F(X) = F(Y)$
H_a: $F(X) \neq F(Y)$

Kolmogorov-Smirnov Test

- Procedure -

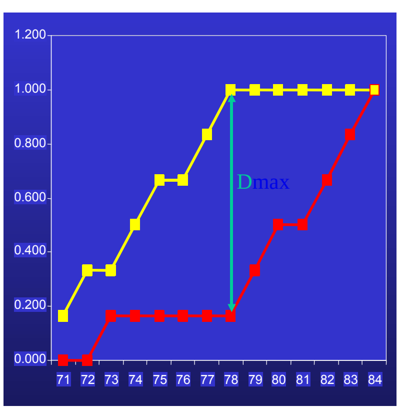
1. Find X_{min} and X_{max} for 2 samples and lay out a column of class categories.
2. List the cumulative frequencies of the two samples in respective columns.
3. Determine relative expected frequencies by dividing by sample sizes.
4. Determine the absolute differences (d) between relative expected frequencies.
5. Identify largest d, becomes D_{max}
6. Multiply D_{max} by n_1n_2 (calc test value).
7. Compare $D_{max}n_1n_2$ with critical value in table.

Kolmogorov-Smirnov Test

- Example: Juniper Data -

Y	Male	Female	M/n1	F/n2	d
71	1	0	0.166	0.000	0.166
72	2	0	0.333	0.000	0.333
73	2	1	0.333	0.166	0.167
74	3	1	0.500	0.166	0.334
75	4	1	0.666	0.166	0.500
76	4	1	0.666	0.166	0.500
77	5	1	0.833	0.166	0.667
78	6	1	1.000	0.166	0.834
79	6	2	1.000	0.333	0.667
80	6	3	1.000	0.500	0.500
81	6	3	1.000	0.500	0.500
82	6	4	1.000	0.666	0.334
83	6	5	1.000	0.833	0.167
84	6	6	1.000	1.000	0.000

← D_{max}



Kolmogorov-Smirnov

Cumulative Expected Frequencies Distribution Plot

Note what D_{max} is evaluating.

Kolmogorov-Smirnov Test

- Example -

In this example, the largest difference is D_{max} = 0.834

$D_{calc} = D_{max} (n_1) (n_2) = 0.834 (6) (6) = 30.02$

$D_{table} = 30$ at $n_1 = 6$, $n_2 = 6$, P = 0.05

$D_{calc} > D_{table}$ therefore, reject H_o (barely)

(NB: decision was closer than MWU test)

Kolmogorov-Smirnov Test

- Using R -

```
> ks.test(male, female)

        Two-sample Kolmogorov-Smirnov test

data:  male and female
D = 0.8333, p-value = 0.02597
alternative hypothesis: two-sided
```

NB: this P-value needs to be multiplied by 2 for a 2-tail test. Thus, *P* = 0.05194 ([exact] same as hand-worked example).

Comparison

Note that we used a constant data set for a reason (only one of these three tests was the "appropriate" test to use).

The consequence of using the incorrect test is an incorrect *P*-value,which is connected to power of test, and ultimately your conclusion. The consequences here were minimal but could be profound.

Kolmogorov-Smirnov:	*P* = 0.05194
Mann-Whitney U-test:	*P* = 0.02597
T-test:	*P* = 0.01645

Wilcoxon Signed-Ranks Test

This is the nonparametric analog to the paired two-sample *t*-test.

It is used in those situations in which the observations are paired and you have not met the assumption of normality.

Assumptions: Differences are continuous
Distribution of differences is symmetric
Differences are mutually independent
Differences all have the same median

Wilcoxon Signed-Ranks Test
- Procedure -

1. Find the difference between pairs.
2. Record the sign of the difference in one column, the absolute value of the difference in the other.
3. Rank the absolute differences from the smallest to the largest.
4. Reattach the signs of differences to the respective ranks to obtain signed ranks, then average to obtain the mean rank.

Wilcoxon Signed-Ranks Test
- Procedure -

Recall that this is really a one-sample test using the differences of the ranks and testing them against μ.

Therefore, for *N* < 20:

$$H_0: E(\bar{r}) = \mu = 0$$

$$H_a: \mu \neq 0;\ \ \mu < 0;\ \ \mu > 0$$

For *N* ≥ 20, remove 1/2 from numerator of *z*.

$$z = \frac{\bar{r} - 1/2 - 0}{\sqrt{(N+1)(2\mathrm{N}_1)/6\mathrm{N}}}$$

Wilcoxon Signed-Ranks Test

- Example -

Heights (cm) for two corn seedlings per pot (*N* = 10 pots).
One seedling treated with electric current, the other not.

Pair	Control	Treated	Sign	Diff	Signed-Rank
1	20.0	14.0	+	6.0	5
2	16.8	15.5	+	1.3	1
3	24.7	14.5	+	10.2	7
4	31.8	7.9	+	23.9	10
5	22.8	19.7	+	3.1	3
6	26.7	19.9	+	6.8	6
7	24.4	25.9	-	1.5	- 2
8	18.2	32.9	-	14.7	- 9
9	16.3	19.6	-	3.3	- 4
10	31.0	42.1	+	11.1	8
				Mean rank	2.5

Wilcoxon Signed-Ranks Test

- Example -

$$z = \frac{\bar{r} - \frac{1}{2} - 0}{\sqrt{(N+1)(2N+1)/6N}} = \frac{2.5 - 0.5}{\sqrt{(10+1)(20+1)/6(10)}} = \frac{2}{\sqrt{3.85}}$$

Z_{calc} = 1.02
The corresponding P-value for Z = 1.02 is 0.154.
Because this is a two-tailed H_o, we multiply by 2.

Thus, P = 0.308, we decide to accept Ho
and conclude that elongation is unaffected by
electric current.

Wilcoxon Signed-Ranks Test

- Using R -

```
> control<-
c(20,16.8,24.7,31.8,22.8,26.7,24.4,18.2,16.3,31.0)
> treated<-
c(14,15.5,14.5,7.9,19.7,19.9,25.9,32.9,19.6,42.1)

> wilcox.test(control, treated, paired=TRUE)

        Wilcoxon signed rank test

data:  control and treated
V = 32, p-value = 0.6953
alternative hypothesis: true location shift is not
equal to 0
```

Tukey-Duckworth Test

Nonparametric statistics provide a useful alternative when assumptions for parametric tests can not be met.

In addition, some techniques are so flexible that only a few critical values have been determined, and the calculations are so simple that they can be performed in your head.

One such “pocket statistic” is the Tukey-Duckworth two-sample test.

Tukey-Duckworth Test
- Procedure -

There is only one assumption to this test and that is that the following inequality is adhered to:

$$4 \leq n_1 \leq n_2 \leq 30$$

H_o: The samples are identical
H_a: The samples are different

The test statistic is *C*.
Exists only as a two-sided test.
There are only two critical values: $C_{0.05} = 7$
$C_{0.01} = 10$

Tukey-Duckworth Test
- Procedure -

1. Determine largest and smallest measurement in each ranked sample.
2. For the sample that contains the largest value of the combined samples, count all measurements that are larger than the largest measurement in the other sample.
3. For the other sample, count all measurements that are smaller than the smallest measurement of the first sample.
4. Add the two counts together (= *C*).

Tukey-Duckworth Test

- Procedure -

Plot-1	Plot-2
80	84
81	89
82	92
83	92
84	92
85	94
86	95
87	96
89	96
92	96
93	98
94	98
96	99
97	101
98	103

Height (cm) of 15 palms randomly sampled from two 10 x 10 m plots in a tropical moist forest.

"Exclusion values"

$C_{calc} = 4 + 3 = 7$ $C_{0.05} = 7$

$C_{calc} \geq C_{0.05}$ therefore reject H_o.

Conclude samples are different.

Largest value

Choosing the Appropriate Two-Sample Statistical Test

Independent Variates	Normal Data, Equal Variances Equal Variance t-test
	Normal Data, Unequal Variances Unequal Variance t-test
	Non-normal Data, Equal Variances Mann-Whitney U-test (Wilcoxon)
	Non-normal Data, Unequal Variances Kolmogorov-Smirnov test
Paired Variates	Normal Data, continuous variable Paired t-test
	Non-normal Data, ranked variable Wilcoxon Signed-Rank test

Doing it in R

Test for assumptions:

`> shapiro.test(A)` [then again for B]	Normality test
`> var.test(A,B)`	Variance test

Parametric 2-sample tests:

`> t.test(A,B)`	2-sample Eq. var. t-test
`> t.test(A,B,var.equal=false)`	Unequal var. t-test

Nonparametric 2-sample tests:

`> wilcox.test(A,B)`	Wilcoxon or MW-U-test
`> ks.test(A,B)`	Kolmogorov-Smirnov

Paired 2-sample tests:

`> t.test(A,B,paired=true)`	Paired t-test
`> wilcox.test(A,B,paired=true)`	Paired Wilcoxon test

The End

Of module 2. Stay tuned for regression and ANOVA.