

## Two sample tests (part II):

What to do if your data are not distributed normally:

Option 1: if your sample size is large enough, don't worry - go ahead and use a t-test (the CLT will take care of non-normal data if your sample size is large enough).

If your data only show minor deviations from normality, you can probably make do with a sample size as small as 15 or 20 (each sample!).

If your data show more serious deviations, then you need a larger sample size before the CLT will take care of things for you. 30 or even 50 might be needed.

This is particularly true if you see long tails in your data.

Option 2: use a different test. One that does not need the normal distribution assumption.

This is your only option if you don't meet the conditions under option 1.

Introducing the Mann-Whitney  $U$  test.

The Mann-Whitney  $U$  test (or Wilcoxon-Mann-Whitney  $U$  test):

The only assumption for this test is random data. You still need to make sure you collected the data randomly.

What are you testing? That the two distributions are the same!

In other words, your  $H_0$  becomes:

$$H_0: D_1 = D_2$$

and

$$H_1: D_1 \neq D_2 \quad (\text{or, of course, "<" or ">"})$$

The MWU test detects differences in distributions. That can mean a number of different things, but what it *does not mean* is that it tests for:

- equal means

or

- equal medians

If you want to test for means or medians, you need to make an additional assumption.

Assume that the distributions are only different in location, but not shape (e.g., two identical binomials, two identical uniforms, etc.)

If you make this assumption, then the MWU test can be used to test for means (or medians).

How safe is this assumption?

Depends:

It's a bit like the equal variance assumption for the  $t$ -test.

You should probably at least look at the distributions for each sample and see if they're approximately the same.

Use histograms or boxplots. You can also use more sophisticated graphics similar to q-q plots, but they're beyond our class.

So your hypotheses are:

$H_0$ : The populations distributions of  $Y_1$  and  $Y_2$  are the same (note capitalized  $Y$ 's).

$H_1$ : The population distributions of  $Y_1$  and  $Y_2$  are not the same.

which was abbreviated as  $H_0: D_1 = D_2$  and  $H_1: D_1 \neq D_2$  above.

OR

You assume equal distributions except for location and do:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Once you've made this choice, you proceed like as for any hypothesis test:

Figure out your  $\alpha$ .

Calculate your test statistic.

As you might guess, this is called " $U$ ", not " $t$ ":

Make your comparison (using  $U$  tables)

Reject or fail to reject.

The complicated bit is calculating your test statistic:

*Important:* if you've had BIOL 214 or 312, the method to calculate  $U^*$  is different. But it eventually gets you the same value for  $U^*$ .

Feel free to do it either way, but I will present the method from our text.

Calculating  $U^*$  by using ranks:

i) Rank all your data. Do both samples at the same time.

You should probably sort your data in each column from smallest to largest before doing this - it'll make things much easier.

ii) Add up the sums of the ranks for each sample. That gives you  $R_1$  and  $R_2$ .

iii) Now we calculate two values for  $U$ :

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

and

$$U' = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

You can also note (if you wish to be a bit lazy) that:

$$U' = n_1 n_2 - U$$

iv) Now pick the larger of  $U$  or  $U'$ , that's your  $U^*$  (i.e.,  $U^* = \max(U, U')$  )

Note that  $R_1 + R_2 = \frac{N(N + 1)}{2}$ , so you can check your work if you calculate both  $R_1$  and  $R_2$ .

(you can also check your work by doing:  $U + U' = n_1 n_2$ )

v) Use this value to look up  $U_{\text{table}}$  using your value of  $\alpha$  (see table B.11, p. 747).

You'll need your sample sizes:

Designate  $n_1$  as the smaller sample and  $n_2$  as the larger sample (it doesn't make any difference to using the table).

If  $U^* \geq U_{\text{table}}$ , reject  $H_0$ ; otherwise fail to reject.

vi) For a one sided test, just use the appropriate row at the top.

(Of course, you need to make sure your data agree with  $H_1$ ).

*Comment:* this table also works if you use the other method of calculating  $U^*$  (as presented in 214/312).

Let's do an example, using 8.11 from the text.

This is presented a bit differently than in your text.

We wish to find out if male and female students are the same height:

$H_0$ : male and female students are the same height

$H_1$ : male and female students are not the same height

*or if we assume equal distributions except for location:*

$H_0$ : mean height of male and female students is the same

$H_1$ : mean height of male and female students is not the same

Let  $\alpha = 0.05$

Now we need to sort and rank our data:

Height of			
rank	males	females	rank
4	170	163	1
6	175	165	2
8	180	168	3
9	183	173	5
10	185	178	7
11	188		
12	193		
Sum	60		18

(Your book does the ranks backwards - giving the highest height a value of 1. It's irrelevant how you do it (your book even mentions this), but I think this way makes more sense).

so now:

$$U = 7 \times 5 + \frac{7(8)}{2} - 60 = 3$$

and

$$U' = 5 \times 7 + \frac{5(6)}{2} - 18 = 32$$

$$(\text{note check: } 60 + 18 = 78 = \frac{12(13)}{2})$$

And  $U^* = \max(3, 32) = 32$ .

Now we need to do our comparison:

For comparison purposes, we let  $n_1 = \text{larger sample size}$  and  $n_2 = \text{smaller sample size}$ , so we have:

$$n_1 = 7$$

$$n_2 = 5$$

Looking at table B.11 and using  $\alpha = 0.05$ , we get  $U_{\text{table}} = 30$ .

Finally:

Since  $U^* = 32 \geq U_{\text{table}} = 30$ , we reject  $H_0$  and conclude that heights of male and female students are different.

(If we had wanted a one sided test (males being taller than females makes sense), we would checked that  $R_1 > R_2$  ( $R_1$  represents the ranks of males, so  $R_1$  should be larger), and then used the row for one sided tests in our table ( $U_{\text{table}} = 29$ )).

Some comments on the Mann-Whitney U test:

The problem of ties:

If two (or more) data points are equal, then take the average of all the ranks and assign that to each data point. For example:

	a	b	
2	3	3	2
2	3	5	4
5	7	9	6

("2" is the average of  $(1 + 2 + 3)/3$ )

Note that the check will still work.

Power of the MWU test:

This test is quite powerful, even compared to the  $t$ -test.

If the data are normal, the  $t$ -test will do better.

If the data are not normal, the MWU test can do much better.

Note that if the data are normal, that does not mean the MWU test is "invalid":

It's just not as good as the  $t$ -test

Why does it work?

We don't have the time to really delve into this. But note the following:

If everything in the first sample is bigger than in the second sample, then  $R_1$  will have the sum of all the bigger ranks, and  $R_1$  will be much bigger than  $R_2$ .

In other words, if  $R_1 \gg R_2$  (" $\gg$ " means "big difference", though it's obviously a bit vague) that implies that the null hypothesis is not true.

On the other hand, if  $R_1 \approx R_2$ , that implies that the ranks are about the same in each sample, and that there really isn't a big difference.

If you want more details, check out Wikipedia.