STATISTICS FOR BUSINESS

TEACHER: JOSEMARI SARASOLA

SPECIAL COVID-19

# VALIDATION: Homogeneity test

Let's suppose we have collected data for men and women. Can we take data altogether to perform a research? That is to say, are men and women homogenous?

To decide about that we perform Wilcoxon's rank.sum test.

Example: califications in a classroom, by sex:

8.2 9.1 7.5 6.8 3.0 5.7 4.4 2.2 1.1 10

w w w  m m w m m m w

Order data and give rank (order level), keeping sex:

1.1 2.2 3.0 4.4 5.7 6.8 7.5 8.2 9.1 10

Ranks= 1 2 3 4 5 6 7 8 9 10

       m m m m w m w w w w

 Wmen=1+2+3+4+6=16

Wwomen=5+7+8+9+10=39

Take Wmin=16

Look tables, alfa=5% eg,: n=5 (corresponding to Wmin), m=5 -→ W* (critical)=19, if Wmin is 19 or lesser reject homogeneity, if it's bigger accept homogeneity

In this case, we reject homogeneity, hence men and women should be analyzed apart from each other.

---

Remark 0: if there are ties, for example 4.4-4.4 (m-w) ranks=6-7, take for both 6.5 as rank value

Remark 1: in the previous example, question is: are men and women different?( so test is two sided). if question is: are women better students? the test is one sided, and we take the one sided table

Remark2: if question is: are men better students? As Wmen>Wwomen, no need to perform the test. The answer is: clearly and absolutely, no.

Remark3: if n, m are outside the table, take normal approx.. set the corresponding normal distribution (look slide 20), with n1 the number of elements giving Wmin, and look for the value in that normal distribution leaving downwards alfa/2 probability.