STATISTICAL HYPOTHESIS TESTING USING NON-PARAMETRIC TESTS

METHODS, INTERPRETATION, AND REAL-LIFE APPLICATIONS

NEGIN HEZARJARIBI

SEPTEMBER 23, 2023

Table of Contents

Introduction

Statistical testing plays a pivotal role in the world of data analysis, enabling us to draw conclusions and validate hypotheses. This paper explores the essential framework of statistical testing, focusing on result interpretation. It introduces two non-parametric tests, the Sign test and the Mann-Whitney U test, as practical tools for assessing dataset differences and correlations. I will provide step-by-step guidelines for these tests and apply the Mann-Whitney U test to investigate age differences between rural-dwelling males and females with hypertension and a history of stroke.

1. General Framework of Statistical Testing and Interpretation of Results

Statistical testing forms the cornerstone of scientific inquiry and data analysis, enabling us to draw population or dataset conclusions from sample data. This framework involves key stages: hypothesis formulation, choose a significance level (α), data collection, test selection, calculation, make a decision and result interpretation. Let's examine these steps individually.

1.1. Hypothesis Formulation

The first step in this framework is formulating hypotheses. In statistical terms, a hypothesis test involves two statements: the 'null hypothesis' (H_0), often an initial assumption about a parameter or population, and the 'alternative hypothesis' (H_1), which represents the researcher's goal and typically contradicts the null hypothesis.

For instance, consider the scenario of evaluating a new website design:

- Null Hypothesis (H_0): The new website design does not result in a higher conversion rate.
- Alternative Hypothesis (H_1): The new website design leads to a higher conversion rate.

Here, the null hypothesis asserts the absence of an effect, while the alternative hypothesis posits the presence of the desired effect.

1.2. Choose a Significance Level (α)

The significance level, denoted as α, holds significant importance in hypothesis testing. It aids in determining the level of evidence required before we reject the null hypothesis. Commonly used values for α include 0.05 or 0.01, signifying a willingness to accept a 5% or 1% risk of committing a Type I error—mistakenly rejecting the null hypothesis when it's actually true.

Additionally, the concept of the "confidence level" ($1-\alpha$) is crucial. Typically, researchers opt for a significance level of 5%, corresponding to a confidence level of 95%. This selection influences the precision of our estimate for the population parameter. A higher confidence level results in wider intervals with lower precision, while a lower confidence level yields narrower intervals, providing a more accurate estimate of the true population value.

1.3. Gathering Data

In data collection, precision and accuracy are paramount. This encompasses methods like observations, experiments, or surveys, chosen for their alignment with the research question and accurate representation of the population or phenomenon. Quality data is essential to faithfully depict the subject under investigation.

1.4. Test Selection

Selecting the appropriate statistical test is pivotal, and it hinges on factors such as data characteristics and research objectives.

Parametric Tests

Parametric tests make assumptions about the population distribution, often assuming normality. They are ideal for exploring relationships between variables. Examples include the t-test for comparing two group means and ANOVA for comparing three or more groups.

Non-Parametric Tests

Non-parametric tests, like the Chi-square, Sign test, and Mann-Whitney U test, offer versatility by not assuming specific data distributions. They become valuable when data diverge from parametric assumptions or involve categorical and ordinal data. These tests excel in small sample sizes and stand strong against distribution assumptions, providing flexibility across various data types.

1.5. Perform the Statistical Test

After selecting the appropriate statistical test, our next step is to apply it to our sample data and compute a test statistic (e.g., t-value, F-value, chi-squared statistic). This test statistic measures how closely our sample data aligns with our initial hypothesis.

In simpler terms, we create a 'test statistic' and establish a 'critical region' based on our chosen significance level (α). If the calculated test statistic, derived from a random sample, falls within this critical region, we reject the null hypothesis. Conversely, if it falls outside this region, we have no strong reason to reject the null hypothesis.

Crucially, we rely on a 'critical value' as our guide in this decision-making process, helping us determine whether to accept or reject the null hypothesis.

1.6. Determine the P-value

The p-value represents the smallest probability of committing a Type I error (significance level) that would result in the rejection of the null hypothesis by the test statistic. Calculating the p-value relies solely on the null hypothesis, without any dependence on the alternative hypothesis. Consequently, decisions regarding the rejection of the null hypothesis (H_0) are based on the calculated p-value.

The p-value is computed according to the type of hypothesis test for the parameter θ as follows:

- When the hypotheses are: H_0: $\theta = \theta_0$ and H1: $\theta < \theta_0$
  P-value = $P(\theta_0\ (X < x))$
- When the hypotheses are: H_0: $\theta = \theta_0$ and H1: $\theta > \theta_0$
  P-value = $P(\theta_0\ (X > x))$
- When the hypotheses are: H_0: $\theta = \theta_0$ and H1: $\theta \neq \theta_0$
  P-value = $2 * \min(P(\theta_0\ (X \leq x)), P(\theta_0\ (X \geq x)))$

1.7. Make a Decision

In this section, we compare the calculated p-value to the chosen significance level (α):

- If the p-value is less than or equal to the chosen significance level (α), we reject the null hypothesis in favor of the alternative hypothesis.
- If the p-value is greater than α, we fail to reject the null hypothesis. In such cases, there is insufficient evidence to support the alternative hypothesis, and we refrain from drawing conclusions against the null hypothesis.

It's important to emphasize that the p-value serves as a measure of evidence against the null hypothesis. However, it should not be interpreted using arbitrary thresholds such as "almost always" or "with high confidence." Instead, its interpretation should align with the chosen significance level (α).

For instance, if the p-value closely approaches α, we should refrain from stating that "almost always, the null hypothesis is rejected." Similarly, if the p-value is significantly smaller than α, we should avoid claiming that "with high confidence, we can reject the null hypothesis."

1.8. Interpreting the Result

Interpreting the results constitutes a critical phase in the hypothesis testing process. Let's explore how to interpret these outcomes:

- Rejecting the Null Hypothesis: When we reject the null hypothesis (p-value $\leq \alpha$), we infer that there exists significant evidence supporting the alternative hypothesis. In simpler terms, the effect, difference, or relationship we hypothesized is likely present within the population.
- Failing to Reject the Null Hypothesis: In cases where we do not reject the null hypothesis (p-value $> \alpha$), it is important to note that we do not conclude that the null hypothesis is true. Instead, we acknowledge that there is insufficient evidence to lend support to the alternative hypothesis.

In addition to these considerations, it is essential to examine the effect size and the practical significance of our findings. When interpreting research results, it is advisable to look beyond numerical values and consider their broader scientific and practical implications within the context and objectives of the study.

2. The Sign Test

The sign test is a non-parametric statistical method employed to assess whether the median of a dataset significantly deviates from a hypothesized value. This test holds particular utility when dealing with small datasets or data that does not conform to a normal distribution. Its applicability extends to various research contexts, including scenarios such as evaluating changes in individuals' weight before and after treatment to ascertain whether an increase, decrease, or stability is evident.

One-Sample Sign Test:

The one-sample sign test finds application when there is a single group of data, and the objective is to assess whether the median (or another pre-determined value) of that dataset significantly diverges from a specific reference value. This test entails a systematic comparison of each data point within the sample to the reference value, with a focus on discerning the sign—whether positive, negative, or zero—of the differences. Subsequently, the counts of positive and negative signs are tallied.

Paired Sign Test (Sign Test for Paired Data):

The paired sign test comes into play when we're working with two related sets of data and want to determine if there's a significant difference between these paired observations. Think of it as a variation of the binomial test.

Here's how it works: we calculate the differences between these paired observations and then look at the signs of these differences—whether they're positive, negative, or zero. We keep track of the counts of positive and negative signs.

In cases where the paired observations have equal values, we call it a 'tie.' In the sign test, we exclude any paired samples with ties and focus on the rest for our analysis.

The Process of the One-Sample Sign Test:

Hypothesis:

- Null Hypothesis (H_0): The median of the dataset is equal to the hypothesized value (typically denoted as "MO").
- Alternative Hypothesis (H_1): The median of the dataset is not equal to the hypothesized value.

Data Collection and Preparation:

Before diving into the sign test, it's essential to ensure that our data is a good fit for this type of analysis. The sign test works well with data that's ordinal or continuous but doesn't follow a normal distribution.

Calculation of the Test Statistic:

For each data point, we calculate the difference between the data point and the hypothesized value (MO). Then, we count the number of positive differences (signs) and negative differences (signs). The test statistic (T) is determined as the smaller of these two counts, with values ranging from 0 to the total number of observations.

Determination of the Critical Value:

To decide whether to reject the null hypothesis, we need to find the critical value that matches our chosen significance level (alpha). This critical value can be obtained from a standard statistical table or by using statistical software.

Making a Decision:

Based on the computed test statistic T:

- If T is less than or equal to the critical value, the appropriate decision is not to reject the null hypothesis (H_0).
- If T is greater than the critical value, the appropriate decision is to reject the null hypothesis (H_0).

Interpreting the Results:

- If we choose to reject the null hypothesis (H_0), it implies that there is substantial evidence to conclude that the median of the dataset deviates from the hypothesized value.
- If we choose not to reject the null hypothesis (H_0), it signifies that there is insufficient evidence to assert that the median of the dataset significantly differs from the hypothesized value.

3. Mann-Whitney U Test

The Mann-Whitney U test, also called the Wilcoxon Rank-Sum test, is a non-parametric statistical approach employed to assess whether there is a significant distinction between two independent groups. Unlike the parametric t-test, which suits continuous data, the Mann-Whitney U test is tailored for data that can be ranked. It proves particularly useful when dealing with datasets that don't conform to the assumptions of a normal distribution. Now, let's delve into its procedure:

Hypotheses:

Null Hypothesis (H_0): There is no significant difference between the two independent groups.

Alternative Hypothesis (H_1): There is a significant difference between the two independent groups.

Data Collection:

Collect and organize the data into two independent groups. Ensure that there is no inherent relationship between the data points within each group.

Rank The Data:

Now, combine the data from both groups and rank all the values from lowest to highest, assigning a rank to each value. In cases where two or more data points share the same value, it is common practice to assign them the average rank.

Calculate the text statistics U:

Next, let's calculate the Mann-Whitney U statistic, denoted as U. In order to perform the Mann-Whitney U test, we utilized the standard formula for U as follows:

$$U = NM + \frac{N(N+1)}{2} - \sum_{xi} Rank(xi) \quad (1)$$

N represents the sample size of one group for which we calculate the test statistic U.

M represents the sample size of the other group.

Now, calculate the test statistics U for both groups:

- u_a : the test statistics U for group1
- u_b : the test statistics U for group2.

The final Mann-Whitney U statistic, denoted as U, is determined as the minimum (u_a , u_b).

Determine the critical value (u_c) and p-value:

To determine the critical value (u_c) and p-value, we have several options. we can use statistical software like R or Python with libraries such as scipy, or commercial software like SPSS, which provides the required the p-value, critical value, and test statistics. Alternatively, we can refer to statistical tables, and for your convenience, I have included a critical value table for the Mann-Whitney U test in the appendix. Although this method is less common nowadays due to the convenience of online calculators and dedicated software tools.

To arrive at a conclusion:

We compare the test statistic U with a critical value. The critical value is determined based on several factors, including the chosen significance level (a), the type of test (whether it's two-tailed or one-tailed), and the sample sizes of both groups. This comparison allows us to make a decision regarding the null hypothesis in our statistical analysis.

- If the test statistic U is less than or equal to critical value, we fail to reject the null hypothesis.
- If the test statistic U is greater than critical value, we reject the null hypothesis.

In hypothesis testing, we compare the p-value to the significance level (a). The significance level (a) represents the probability of making a Type I error, which is the error of incorrectly rejecting a correct null hypothesis (H_0). It sets the threshold for statistical significance. When the p-value is larger than the significance level, it suggests that the chance of making a Type I error, wrongly rejecting H_0, is relatively high. A larger p-value provides more support for H_0 or the null hypothesis.

Interpreting the result:

- If we fail to reject the null hypothesis (H_0), it suggests that there is not enough evidence to conclude a significant difference between the two groups.
- If we reject the null hypothesis (H_0), it indicates that there is enough evidence to support a significant difference between the two groups.

4.Real-life Example For Mann-Whitney U Test:

In this section, we will conduct a Mann-Whitney U test using a real-life example. Our objective is to determine whether a significant difference exists in the ages of male and female patients living in rural areas, diagnosed with hypertension, and who have experienced a stroke. To address this question, we have selected a subset of 'The Stroke Prediction Dataset,' which was provided by FEDESORIANO on Kaggle. This dataset consists of 11 clinical features used to predict stroke events (FEDESORIANO, n.d., retrieved 2023-09-14).

We present the ages of females and males who meet the criteria mentioned above:

Female: 79, 81, 50, 82, 72, 39, 80, 67, 80, 67, 66, 50, 73, 68, 57, 81, 79, 80

Male: 74, 76, 71, 61, 56, 69, 76, 79, 72, 78, 68, 70, 78, 71

In this analysis, the use of the Mann-Whitney U test is justified as it accommodates real-world data that may deviate from normality, making it suitable for comparing the ages of male and female patients. This non-parametric test is a robust alternative when parametric assumptions cannot be met.

For this analysis, we will consider a significance level of $\alpha = 0.05$ and employ a two-tailed test. This choice is made because the ages can differ in either direction, whether males are older or younger than females. This approach is conservative as it accommodates differences in both directions.

We have formulated our null hypothesis (H0) asserting that no significant difference exists between the ages of females and males living in rural areas with hypertension who have had a stroke. Our alternative hypothesis (H1) posits a significant difference in ages between these groups.

Next, we will combine the data from both groups, rank all the values from lowest to highest, and assign ranks to each value. In cases where two or more data points share the same value, they are typically assigned the average rank.

39(1), 50(2.5), 50(2.5), 56(4), 57(5), 61(6), 66(7), 67(8.5), 67(8.5), 68(10.5), 68(10.5), 69(12), 70(13), 71(14.5), 71(14.5), 72(16.5), 72(16.5), 73(18), 74(19), 76(20.5), 76(20.5), 78(22.5), 78(22.5), 79(25), 79(25), 79(25), 80(28), 80(28), 80(28), 81(30.5), 81(30.5), 82(32).

Here are the ranks alongside the ages on the left-hand side:

Female: 79 (25), 81 (30.5), 50 (2.5), 82 (32), 72 (16.5), 39 (1), 80 (28), 67 (8.5), 80 (28), 67 (8.5), 66 (7), 50 (2.5), 73 (18), 68 (10.5), 57 (5), 81 (30.5), 79 (25), 80 (28)

Male: 74 (19), 76 (20.5), 71 (14.5), 61 (6), 56 (4), 69 (12), 76 (20.5), 79 (25), 72 (16.5), 78 (22.5), 68 (10.5), 70 (13), 78 (22.5), 71 (14.5)

Our next task is to sum all these ranks for each group:

Sum of Ranks for Females (T_f) = $\sum_{xi} Rank(xi)$ =
25+30.5+2.5+32+16.5+1++28+8.5+28+8.5+7+2.5+18+10.5+5+30.5+25+28= 307

Sum of Ranks for Males (T_m): $\sum_{xi} Rank(xi)$=
19+20.5+14.5+6+4+12+20.5+25+16.5+22.5+10.5+13+22.5+14.5=221

Now, we are ready to perform the calculation for the U statistic test. The formula is as follows:

$$U = NM + \frac{N(N+1)}{2} - \sum_{xi} Rank(xi) \quad (1)$$

We will calculate for females:

N =18 , M= 14 , $\sum_{xi} Rank(xi)$ = T_f =307, then: $u_f = 18 * 14 + \frac{18(18+1)}{2} - 307 = 116$

we will calculate for males:

N= 14 , M= 18 , $\sum_{xi} Rank(xi)$ =T_m =221, then: $u_m = 14 * 18 + \frac{14(14+1)}{2} - 221 = 136$

the Mann-Whitney U statistic = $u_{statistic}$ =min (u_f , u_m)= u_f =116

After calculating the Mann-Whitney U statistic, we consulted the critical value table included in the appendix (Table 1) to determine the critical value (u_c) for our two-tailed test. With sample sizes N1 = 18 and N2 = 14, we identified a critical value of $u_c = 74$. Comparing this critical value to our computed test statistic ($u_{statistic} = 116$), we arrived at our decision regarding the null hypothesis.

We then calculate the p-value, which is 0.7178, exceeding our chosen significance level (alpha = 0.05). The p-value was computed using Python with libraries like scipy, commonly employed for statistical analyses.

Interpretation:

If the test statistic were greater than the critical value, we would typically reject the null hypothesis (H_0). However, in this case, the p-value (0.7178) exceeds the chosen significance level (alpha = 0.05), indicating that the probability of making a Type I error, which involves incorrectly rejecting a correct null hypothesis (H_0), is too high at 71.78%. A larger p-value provides stronger support for H_0. In summary, we fail to reject the null hypothesis, indicating insufficient evidence to assert a significant age difference between females and males in rural areas with hypertension and have experienced a stroke.

Conclusion

In summary, this paper has explored the vital role of statistical testing in data-driven inferences, emphasizing a systematic approach. We've introduced two non-parametric tests, the Sign test and the Mann-Whitney U test, as valuable tools for analyzing data under non-normality assumptions. Through comprehensive guidance on conducting and interpreting these tests, we've applied the Mann-Whitney U test in a practical scenario, equipping readers with a solid understanding of statistical testing, the Sign test, and the Mann-Whitney U test for confident real-world data analysis.

References

1. FEDESORIANO. (n.d.). *Stroke Prediction Dataset*. Kaggle. Retrieved September 14, 2023, from https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

2. Mirzaei, K. (2009-2010). پژوهش، پژوهشگری و پژوهشنامه نویسی [*Research, Researcher, and Writing a Research Paper*]. انتشارات جامعه شناسان.

3. Contributing Authors. (2014). *Critical Values for the Mann-Whitney U Test*. UMass Boston OCW. https://ocw.umb.edu/psychology/psych-270/other-materials/RelativeResourceManager.pdf

4. Shier, R. (2004). *Statistics: 2.1 The Sign Test*. Mathematics Learning Support Centre. Retrieved from https://www.statstutor.ac.uk/resources/uploaded/signtest.pdf

5. Sirisilla, S. (2022, June 10). *6 Steps to Evaluate the Effectiveness of Statistical Hypothesis Testing*. Enago Academy. https://www.enago.com/academy/evaluate-statistical-hypothesis-testing/

6. Kraska-Miller, M. (2014). *Nonparametric Statistics for Social and Behavioral Sciences*. CRC Press, Taylor & Francis Group.

7. Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

Appendix

Critical Values of the Mann-Whitney U

Table1. Critical Values of the Mann-Whitney U (Two-Tailed Testing)

n_2	α	n_1																	
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	0	0	1	2	2	3	4	4	5	5	6	7	7	8	9	9	10	11
	.01	--	0	0	0	0	0	1	1	1	2	2	2	3	3	4	4	4	5
4	.05	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
	.01	--	--	0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	10
5	.05	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
	.01	--	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6	.05	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
	.01	--	1	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22
7	.05	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
	.01	0	1	3	4	6	7	9	11	12	14	16	17	19	21	23	24	26	28
8	.05	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
	.01	0	2	4	6	7	9	11	13	15	17	20	22	24	26	28	30	32	34
9	.05	4	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
	.01	1	3	5	7	9	11	14	16	18	21	23	26	28	31	33	36	38	40
10	.05	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
	.01	1	3	6	8	11	13	16	19	22	24	27	30	33	36	38	41	44	47
11	.05	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
	.01	1	4	7	9	12	15	18	22	25	28	31	34	37	41	44	47	50	53
12	.05	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
	.01	2	5	8	11	14	17	21	24	28	31	35	38	42	46	49	53	56	60
13	.05	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
	.01	2	5	9	12	16	20	23	27	31	35	39	43	47	51	55	59	63	67
14	.05	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
	.01	2	6	10	13	17	22	26	30	34	38	43	47	51	56	60	65	69	73
15	.05	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
	.01	3	7	11	15	19	24	28	33	37	42	47	51	56	61	66	70	75	80
16	.05	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
	.01	3	7	12	16	21	26	31	36	41	46	51	56	61	66	71	76	82	87
17	.05	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
	.01	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77	82	88	93
18	.05	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
	.01	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88	94	100
19	.05	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
	.01	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	107
20	.05	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
	.01	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114

Table2. Critical Values of the Mann-Whitney U (One-Tailed Testing)

n_2	α	n_1																	
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	3	3
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	.05	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	.01	--	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	.05	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
	.01	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	.05	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
	.01	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	.05	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
	.01	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12	.05	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
	.01	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	.05	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
	.01	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	.05	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
	.01	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	.05	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
	.01	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	.05	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
	.01	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	.05	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
	.01	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	.05	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
	.01	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	.05	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
	.01	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	.05	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
	.01	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

Note: These tables are adapted from 'Critical Values for the Mann-Whitney U Test' from UMass Boston OCW(https://ocw.umb.edu/psychology/psych-270/other-materials/RelativeResourceManager.pdf),