Using Statistics to Test Hypotheses

How statistics can be used to test an hypothesis – an example:

Imagine that you are trying to prove the hypothesis that "people from Monroe have smellier breath than people from Snohomish. You are comparing breath smell readings from a smell meter (0 = no smell; 10 = Extreme Smell). Your first group of test subjects, made up of citizens from Snohomish, has breath scores of 7, 3.5, 5.5, 8, 6.5, 4, and 6.5. Your second group, from Monroe, has breath scores of 6, 10, 7.5, 4.5, 8, 7, and 9. After averaging your subjects' breath scores, you get 5.9 for the Snohomish group and 7.4 for the Monroe group.

The big question here is what you should conclude from your study. Remember that you only measured the breaths of six individuals from each community. If you assume that your measurements were precise and accurate, and that your test subjects were chosen randomly, is your measured difference in breath scores significant enough for you to confidently say that "people from Monroe have smellier breath?" Within every population, there is natural variation. Given your results, is it possible that there is not a real difference between the breaths of people in Snohomish and people in Monroe? Is it possible that, purely by chance, you happened to randomly select stinkier than average Monroe citizens – or that you happened to randomly select Snohomish citizens who were less stinky than average? Of course it is. But how likely is it? What are the chances of randomly selecting groups that are as different as your test groups *if there really is no difference* between the breaths of people from Snohomish and Monroe? If the chance of accidentally selecting groups this different (from populations with no real difference) is very low, then there probably is a real difference between the groups.

How do you find out if that chance is, in fact, very low? You use statistics. You use a statistical test to determine the exact probability of the difference observed in your results, given no real difference between the populations. Then you can say something like, "the chance of randomly selecting groups of people that are this different, *from two populations that have no real difference*, is 0.01 – or, in other words, a one in one hundred chance. There is only a 1% chance that people from Monroe do not actually have smellier breath. Therefore, it is very likely that the citizens of Monroe really do have smellier breath."

Hopefully, this example illustrates that you're not using statistics to directly prove that your hypothesis is very likely. What you're doing is using statistics to show that it is very unlikely that your hypothesis is wrong.

Alternate Hypothesis: Though it can have a slightly different meaning in other situations, in this class, **your alternative hypothesis states that there really is a difference between the two groups that you are comparing.** For instance, you might be trying to show that students who have an SSR class *really do* read better than students who don't have an SSR class – or that papers glued with glue stick really do hold together with more force than papers that are glued together by slug slime.

Null Hypothesis: This hypothesis states that there is no real difference between the two groups you are comparing – that any *observed* difference between two groups is the result of chance and natural variation within the groups. You construct a null hypothesis for the sole purpose of showing that it is unlikely, and that your alternate hypothesis is, therefore, likely. It's not exactly the *opposite* of your alternate hypothesis, but it's close.

P-Value: The probability of something, expressed as a decimal.

Significance Cutoff: The maximum p-value for the null hypothesis in order for the results of a study to be "significant." The symbol for the significance cutoff is α. The most commonly accepted p-value for scientific studies is 0.05. In other words, in order for your results to be "significant," there needs to be a less than 5% chance that your hypothesis is wrong.

α = alpha

One-Sided and Two-Sided Null Hypotheses (also called one and two-tailed):

- Example of a simplified one-sided null hypothesis: "Men are not taller than women."
- Example of a simplified two-sided null hypothesis: "Men are neither shorter nor taller than women."

The Mann-Whitney Test:

This is the statistical test that we will be using in this class. We will be using it because it is relatively simple. It is often called a "rank sum" test, because all you have to do is rank things and add them up.

Keep in mind that this is just one of many statistical tests. Professionals pick the statistical test that best suits their particular investigation. We will be doing the opposite; you will need to pick an investigation that suits this test, since this is the only one you will know. If you would like to do some self-teaching, feel free to use some other statistical test, but if you do that, you're on your own.

One Sided Mann-Whitney Test: The example below shows how to test the hypothesis: Men are not taller than women.

1) Collect data from two groups. For example, heights of men and heights of women.
2)
   a) Put the values for <u>all</u> (men and women) of the samples together, and assign each sample (person) a rank. The lowest value has a rank of 1. The next lowest has a rank of 2…
   b) What to do if there is a tie? Imagine that four people have these heights (in cm) ⇒ 150, 155, 155, 160. The 150 cm person has a rank of 1. The two 155 cm people share 2nd and 3rd place, so they each get a rank of 2.5 (in the middle). The 160 cm person gets a rank of 4.
3) Get the "rank sum" for the group which you expect to have the lower values (the women, because we are expecting them to be shorter). To get the rank sum, add up all of the rankings for samples in that group (add up all of the women's ranks).
4)
   a) Use a table to find the "critical rank sum" for the α (significance cutoff) you have chosen [In this class, we will use p=.05].
   b) The letter "n" stands for sample size. You do not have to have the same sample size for men and women. However, when you read the table, the "y axis" of the table corresponds to the group for which you have more samples, and the "x axis" corresponds to the group with ~~less~~ fewer samples. If you have an equal number of samples, it makes no difference. <u>Find the critical rank sum by going down the chart to the correct sample size for the more numerous group and by going across to the sample size of the less numerous group.</u>
   c) Values are listed for four different significant cut-offs (α). Make sure you are using 0.05.
   d) Make sure you are looking at α for a one-sided test.
5) If the rank sum for the group you expected to have lower values (women) is less than or equal to the rank sum, the results are significant at that α (95% confidence).

Two-sided Mann-Whitney Test: If you wanted to test the hypothesis that men are neither shorter nor taller than women, you would calculate a rank sum for both groups. Then you would use the lower rank sum and compare that value to the critical rank sum for a two-sided test.

Hypotheses + Data

① People from Bothell have smellier breath.
Breath Scores:
Bothell: 6, 10, 7.5, 4.5, 8, 7, 9
Edmonds: 7, 3.5, 5.5, 8, 6.5, 4, 6.5 (10 = stinkiest)

② Female cats have shorter tails.
Lengths (inches):
Males: 25, 29, 34, 20, 18, 27
Females: 17, 25, 21, 19, 24

③ People over 40 feel happier than people under 40.
Happiness (10=most)
Under 40: 8, 1, 3, 5, 6, 7, 6
Over 40: 10, 1, 4, 7, 7, 10, 9

The table below was taken from...

Rice, John A. (1988). *Mathematical statistics and data analysis.* Wadsworth, Inc.

Table 8 Critical values of smaller rank sum for the Wilcoxon Mann–Whitney test

n_2	α for Two-Sided Test	α for One-Sided Test	n_1 (Smaller Sample)																			
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.20	.10		3	7																	
	.10	.05			6																	
	.05	.025																				
	.01	.005																				
4	.20	.10		3	7	13																
	.10	.05			6	11																
	.05	.025				10																
	.01	.005																				
5	.20	.10		4	8	14	20															
	.10	.05		3	7	12	19															
	.05	.025			6	11	17															
	.01	.005					15															
6	.20	.10		4	9	15	22	30														
	.10	.05		3	8	13	20	28														
	.05	.025			7	12	18	26														
	.01	.005				10	16	23														
7	.20	.10		4	10	16	23	32	41													
	.10	.05		3	8	14	21	29	39													
	.05	.025			7	13	20	27	36													
	.01	.005				10	16	24	32													
8	.20	.10		5	11	17	25	34	44	55												
	.10	.05		4	9	15	23	31	41	51												
	.05	.025		3	8	14	21	29	38	49												
	.01	.005				11	17	25	34	43												
9	.20	.10	1	5	11	19	27	36	*46	58	70											
	.10	.05		4	*10	16	24	33	43	54	66											
	.05	.025		3	8	14	22	31	40	51	62											
	.01	.005			6	11	18	26	35	45	56											
10	.20	.10	1	6	12	20	28	38	49	60	73	87										
	.10	.05		4	10	17	26	35	45	56	69	82										
	.05	.025		3	9	15	23	32	42	53	65	78										
	.01	.005			6	12	19	27	37	47	58	71										
11	.20	.10	1	6	13	21	30	40	51	63	76	91	106									
	.10	.05		4	11	18	27	37	47	59	72	86	100									
	.05	.025		3	9	16	24	34	44	55	68	81	96									
	.01	.005			6	12	20	28	38	49	61	73	87									
12	.20	.10	1	7	14	22	32	42	54	66	80	94	110	127								
	.10	.05		5	11	19	28	38	49	62	75	89	104	120								
	.05	.025		4	10	17	26	35	46	58	71	84	99	115								
	.01	.005			7	13	21	30	40	51	63	76	90	105								
13	.20	.10	1	7	15	23	33	44	56	69	83	98	114	131	149							
	.10	.05		5	12	20	30	40	52	64	78	92	108	125	142							
	.05	.025		4	10	18	27	37	48	60	73	88	103	119	136							
	.01	.005			7	*13	22	31	41	53	65	79	93	109	125							
14	.20	.10	1	*8	16	25	35	46	59	72	86	102	118	136	154	174						
	.10	.05		*6	13	21	31	42	54	67	81	96	112	129	147	166						
	.05	.025		4	11	19	28	38	50	62	76	91	106	123	141	160						
	.01	.005			7	14	22	32	43	54	67	81	96	112	129	147						
15	.20	.10	1	8	16	26	37	48	61	75	90	106	123	141	159	179	200					
	.10	.05		6	13	22	33	44	56	69	84	99	116	133	152	171	192					
	.05	.025		4	11	20	29	40	52	65	79	94	110	127	145	164	184					
	.01	.005			8	15	23	33	44	56	69	84	99	115	133	151	171					
16	.20	.10	1	8	17	27	38	50	64	78	93	109	127	145	165	185	206	229				
	.10	.05		6	14	24	34	46	58	72	87	103	120	138	156	176	197	219				
	.05	.025		4	12	21	30	42	54	67	82	97	113	131	150	169	190	211				
	.01	.005			8	15	24	34	46	58	72	86	102	119	136	155	175	196				