Nonparametric Statistics

Nonparametric Statistics

Nonparametric statistics refer to a statistical method wherein the **data is not required to fit a normal distribution**.

Nonparametric statistics uses data that is often **ordinal**, meaning it does not rely on numbers, but rather a ranking.

Nonparametric Statistics

Unlike parametric statistics, nonparametric statistics make no assumptions about the probability distributions of the variables being assessed

Nonparametric tests are sometimes called distribution-free tests

Nonparametric Statistics

Nonparametric methods can work with **Ordinal data**

Nonparametric methods convert Measurement data into Signs, Ranks, Signed Ranks, or Rank Sums

Nonparametric methods work better in dealing with Outliers, Skewed data, and Small Samples

NP methods have less Power

Nonparametric Statistics

- One reason that NP methods can use data from any Distribution is that they usually don’t work directly with the data
- **The Sample data are converted to Signs and/or Ranks and the numerical values of the data are lost before any calculations are done**
- So, it makes no difference what type of distribution the source data have

Nonparametric Statistics

- Nonparametric methods work with **Medians** instead of Means

- The Median has advantages over the Mean:
  - It is less influenced by Outliers
  - It is less influenced by the Skew in Skewed data

Nonparametric Statistics

Signs

Specified Value = 30										
Sample Data	28	31	30	33	32	28	30	31	27	32
Sign	−	+	0	+	+	−	0	+	−	+

Nonparametric Statistics

Ranks

Sample Data	27	28	28	30	30	31	31	32	32	33
Rank	1	2.5	2.5	4.5	4.5	6.5	6.5	8.5	8.5	10

Nonparametric Statistics

Signed Ranks

Combine the concepts of Signs and Ranks

Sample data	–6	–2	0	+4	+7	+8	+11	+12	+14	+16
1. Sign	0	0	0	1	1	1	1	1	1	1
2. Absolute Value	6	2	0	4	7	8	11	12	14	16
3. Absolute Rank	4	2	1	3	5	6	7	8	9	10
4. Signed Rank	0	0	0	3	5	6	7	8	9	10

Nonparametric Statistics

- Signed Rank tests are the NP counterpart to the Dependent Samples (Paired Samples) t-test

- Rank Sum tests are the NP counterpart the Independent Samples (2-Samples) t-test

Nonparametric Statistics

- **Rank Sums**
  - Comparing Samples taken from two Independent Populations
  - The data values of one Population are not influenced by the data values of another

Nonparametric Statistics

- Step 1: Group: Put all the data from both Samples into a single group (but keep track of which ones came from which group)

- Step 2: Rank: Rank the values in the combined group.

- Step 3: Rank Sum: Total the Ranks for each Sample

Nonparametric Statistics

1. Data, Sample 1	12	7	15	13	13					
1. Data, Sample 2						11	16	8	6	12
2. Ranks	5.5	2	9	7.5	7.5	4	10	3	1	5.5
3. Rank Sum	31.5					26.5				

Nonparametric Statistics

Nonparametric Test	What it does	Parametric Counterpart
Wilcoxon Signed Rank	Compares 1 Median to a specified value	*z*-test, 1-Sample *t*-test
	Compares 2 Dependent (Paired) Medians	Paired (Dependent) Samples *t*-test
Mann–Whitney	Compares 2 Independent Medians	2 (Independent) Samples *t*-test
Kruskal–Wallis	Compares 3 or more Medians, 1 Variable	1-way ANOVA

Nonparametric Statistics

There are some situations when it is clear that the outcome does not follow a normal distribution

- Using an **Ordinal Scale**
- When the outcome is a rank
- When there are outliers

Nonparametric Statistics

Hypothesis Testing in nonparametric statistics

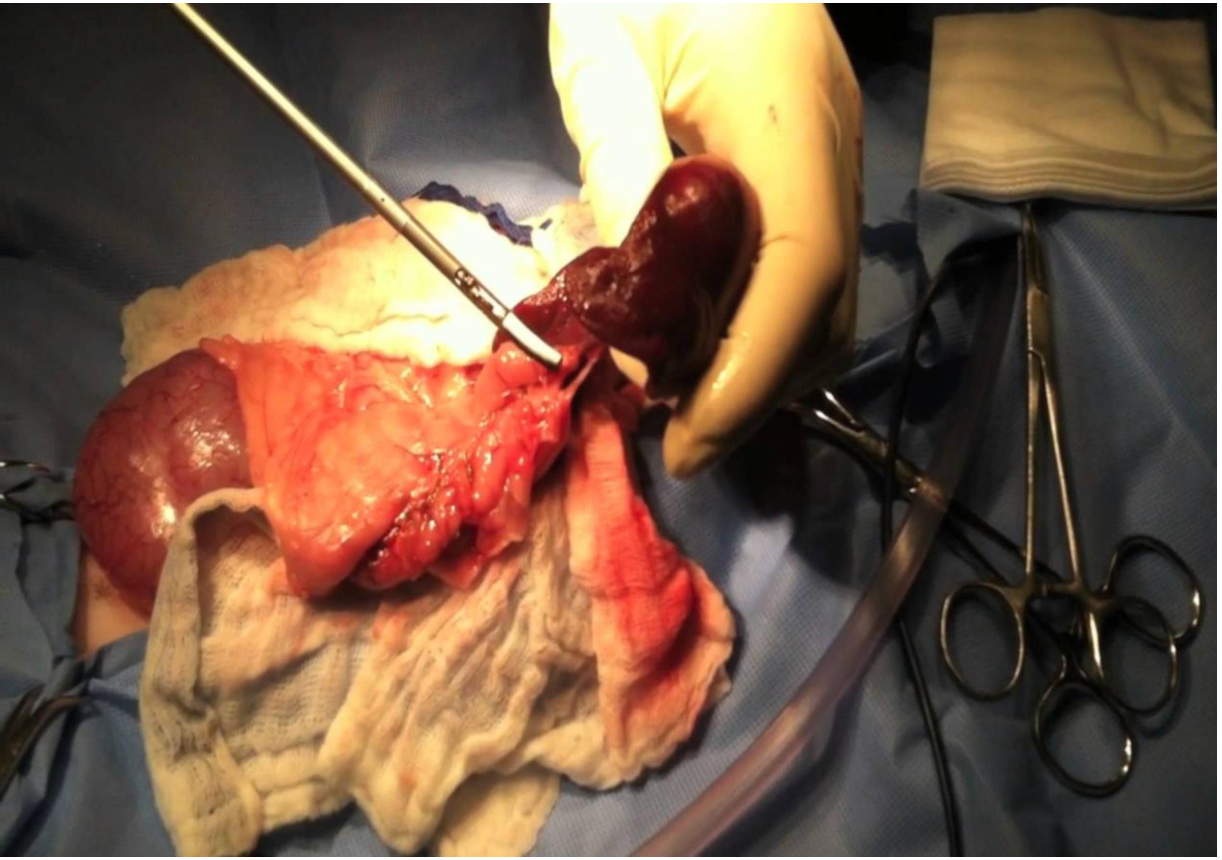
The steps of the hypothesis test are the same as for the t-test, but the null hypothesis is **related to the median rather than the mean**

Example

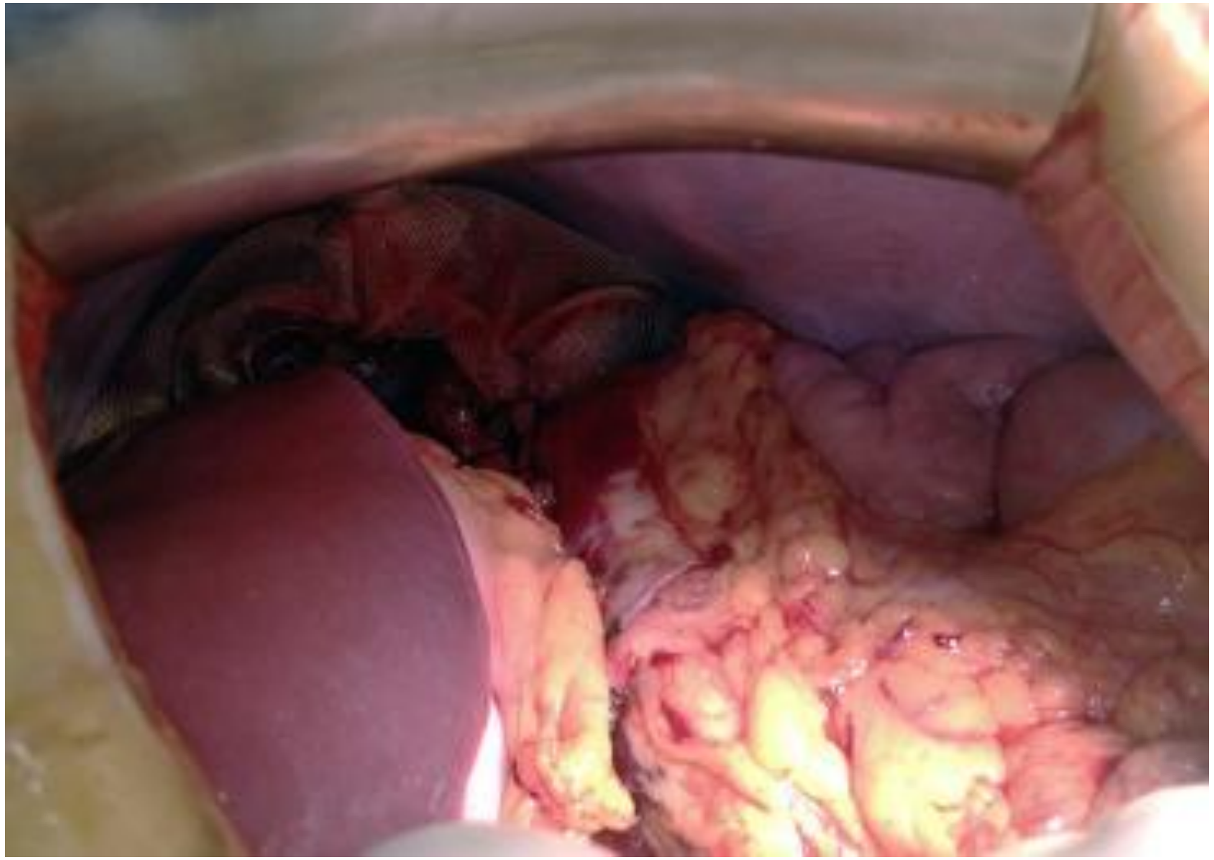
- When patients have pancreatic cancer, often surgery is required to remove the part of the pancreas that has the cancer. When these surgeries are completed, the surgeon has the option to do a more complex surgery to preserve the spleen (splenic preservation) or to remove the spleen as part of the surgery (splenectomy)

- A study was done to compare the two surgical options in terms of health outcomes, cost and time burden on surgical staff

Splenectomy



Spleen Preserved



Question 1

- A question for each technique is to determine the effect of the surgery on the platelet count in patients. Platelets are involved in clotting of patients and patients in surgery are sometimes given drugs to limit the amount of clotting during surgery. A large change in the number of platelets can be a sign that the surgery was particularly difficult.

- For each technique, the surgeons wanted to determine if there is a significant difference in the pre- and post-surgery platelet count.

Example

- First, we will look at the splenic preservation group
- Note that we have paired observations on each of the patients
- We are interested in the difference between the two measurements
- Does it appear there is a difference?

Patient	Pre	Post	Diff
1	260	223	37
2	216	149	67
3	427	224	203
4	217	181	36
5	613	708	-95
6	245	197	48
7	371	303	68
8	236	168	68
9	421	312	109
10	677	521	156
11	363	202	161

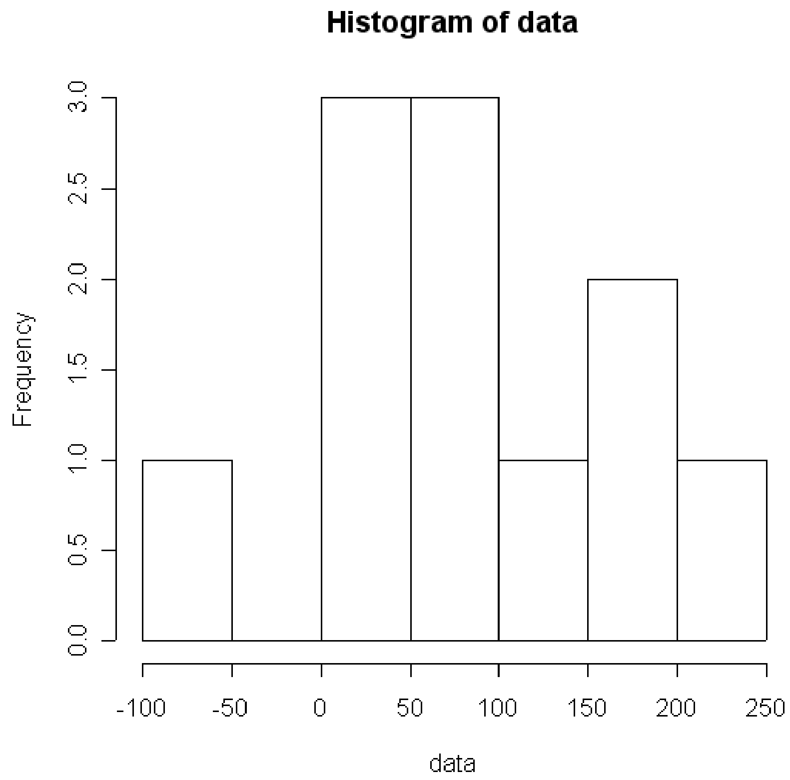
Nonparametric Statistics

- Since we have paired data, we could use the paired t-test.

- What can you say about the distribution of the differences?

- Does the normality assumption of the paired t-test seem appropriate?

- The difference in platelet count may be variable and contain outliers



Nonparametric Statistics

- The null hypothesis for our investigation is that there is no difference in the platelet count before and after the surgery.
- For the two-sample t-test, this was written as
  - H_0: mean difference (pre-post) is equal to zero (δ=0)
- In this case, we have outliers, so the mean is not a good measure of central tendency.
- What measure do you think we should use instead?
- How can we set up and test the appropriate null hypothesis?

Sign test

- The simplest nonparametric test is the sign test

- The null and alternative hypothesis for the sign test
  - H_0 : median of differences (pre-post) = 0
  - H_A : median of differences (pre-post) ≠ 0

- Under the null hypothesis, we would expect the same number of positive and negative signs. Therefore, P(positive sign)=0.5 under the null hypothesis

- If most or all of the differences are positive, there would be some evidence against the null hypothesis.

Nonparametric Statistics

- We have now included the sign column
- If there was truly no effect of the therapy, we would assume that there would be an equal number of + and - signs
- What can you see about the signs of the differences? Is there a significant difference between the two groups? How can we calculate the p-value?

Patient	Pre	Post	Diff	Sign
1	260	223	37	+
2	216	149	67	+
3	427	224	203	+
4	217	181	36	+
5	613	708	-95	-
6	245	197	48	+
7	371	303	68	+
8	236	168	68	+
9	421	312	109	+
10	677	521	156	+
11	218	202	16	+

Remember that a p-value is the probability of obtaining the observed value or something more extreme under the null hypothesis (p=0.5). For the sign test, this is the probability of the observed number of positive signs or more. To make the test two sided, we must take into account the values this extreme from the other side.

$$P(10+'s) = \binom{11}{10}(0.5)^{10}(0.5)^{1} = 0.0054$$

$$P(11+'s) = \binom{11}{11}(0.5)^{11}(0.5)^{0} = 0.00049$$

$$P(0+,1+,10+,11+) = 0.0117$$

Hypothesis test

1) Paired data, alpha level=0.05
2) Hypotheses
   - H_0: median of differences = 0
   - H_A: median of differences ≠ 0
3) Test statistic is 10+ signs
4) **Asymptotic probability** = 0.0117
5) Reject null hypothesis
6) Conclusion: There is a significant difference between the pre- and post-surgery platelet values for patients who had the splenic preservation surgery

Example

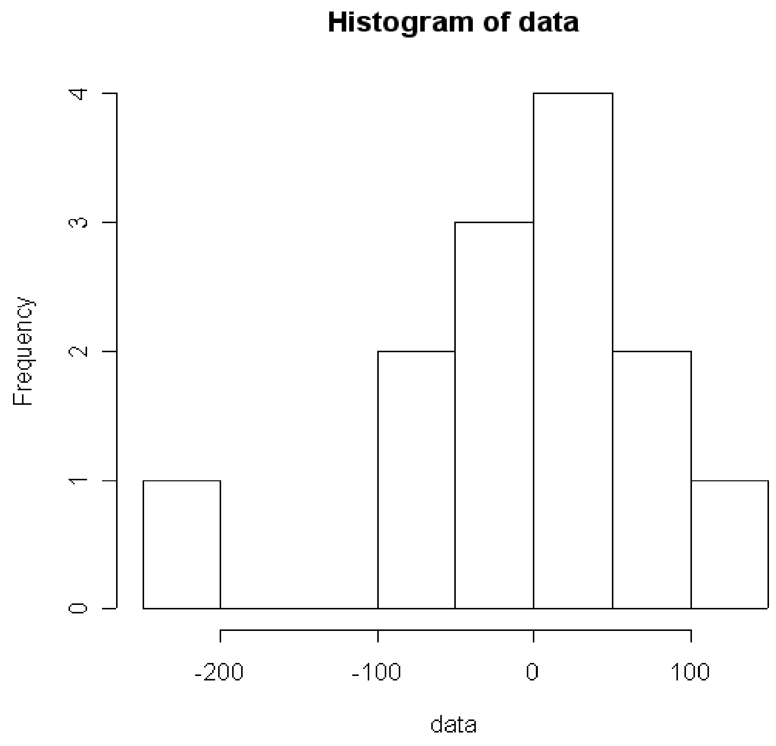
- Now, we can look at the splenectomy group
- Again, we have paired observations on each of the patients, and we are interested in the difference between the two measurements
- Does it appear there is a difference?

Patient	Pre	Post	Diff
1	492	375	117
2	297	382	-85
3	272	325	-53
4	367	585	-218
5	206	181	25
6	284	237	47
7	338	273	65
8	212	243	-31
9	161	147	14
10	384	326	58
11	224	214	10
12	251	292	-41
13	224	263	-39

Nonparametric Statistics

- Again, the distribution of the differences does not appear normal

- We could use the sign test, but there is another more powerful test called the Wilcoxon rank sum test



Wilcoxon signed rank

- The sign test looks only at the sign of the differences, but the Wilcoxon signed rank uses the sign and rank of the differences.

- The null and alternative hypotheses are the same as for the sign test
  - H_0: median diff = 0
  - H_A: median diff ≠ 0

Patient	Pre	Post	Diff	Rank
1	492	375	117	12
2	297	382	-85	-11
3	272	325	-53	-8
4	367	585	-218	-13
5	206	181	25	3
6	284	237	47	7
7	338	273	65	10
8	212	243	-31	-4
9	161	147	14	2
10	384	326	58	9
11	224	214	10	1
12	251	292	-41	-6
13	224	263	-39	-5

- The test statistic of this test is the sum of the positive ranks.
- **Under the null hypothesis, half of the ranks should be positive and half of the ranks should be negative.**
- Evidence against the null would be having the sum of the positive ranks either being very high or very low.

Hypothesis test

- Paired data, Wilcoxon test, alpha=0.05
- Hypotheses
  - Null: median difference = 0
  - Alternative: median difference not = 0
- Test statistic: Sum of positive ranks = 44
- **Asymptotic probability** = 0.946
- Fail to reject null hypothesis

- **Conclusion: There is no enough significant statistical evidence of a difference between the pre- and post- platelet counts for patients who had a splenectomy during their surgery.**

Conclusions

- Our hypothesis tests show that patients from the splenic preservation group have a significant change in their platelet count after surgery ($p=0.01$) and patients from the splenectomy group do not have a significant change ($p=0.94$). These results may show that the splenic preservation surgery is difficult on the patient and other measures should be investigated to ensure that this surgery is not overly stressful on patient systems.

- For the actual study several other markers were investigated because platelets only tells a small part of the story.

Comments

- When we have paired data and the assumptions of a paired t-test are not met, we have two ways to complete the hypothesis test

- **The Wilcoxon test is always preferred over the sign test because it uses more of the data (since it uses the ranks). The Wilcoxon test has much more power to detect a significant difference.**

- There is not a large loss of power in using a Wilcoxon test compared to a t-test when the normality assumption holds. The Wilcoxon is much more powerful when the normality assumption does not hold.

- Therefore, the Wilcoxon test is more appropriate if there is any reason to doubt the normality assumption.

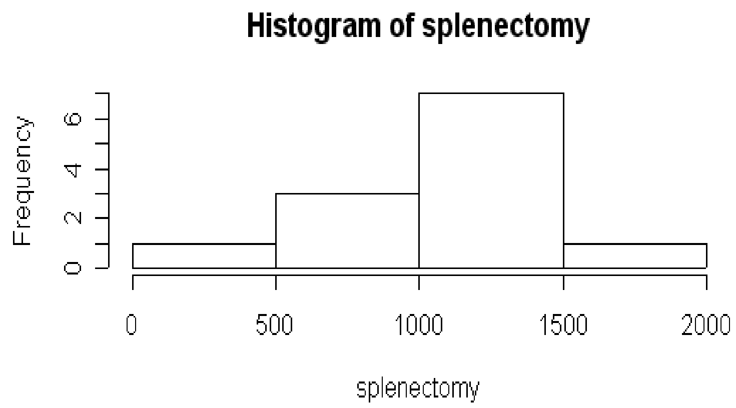
Question 2

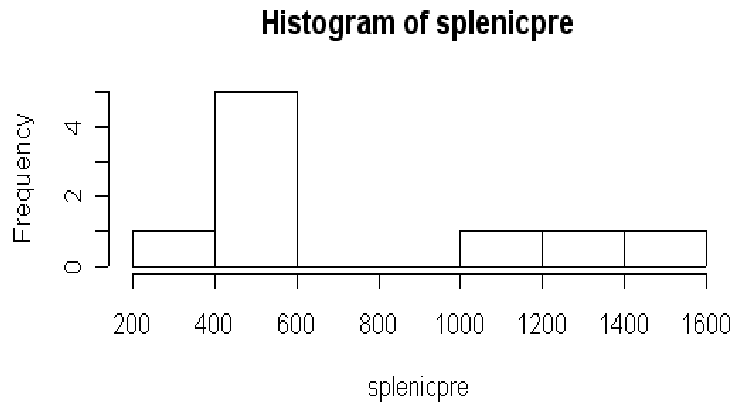
- Beyond the surgical outcomes, the surgeons were also interested in the economics of the two types of surgery.
- One of the costs of interest is the anesthesia cost. The cost (in dollars) for several of the patients in each of the two groups is given here

Splenectomy	**Splenic preservation**
1011.07	496.44
1066.82	541.76
610.80	1562.01
1111.44	2515.12
955.68	1133.99
1203.84	300.33
1600.32	482.55
555.90	503.22
1302.95	2744.23
182.34	1232.22
1233.20	
1402.09	

Nonparametric Statistics

- We want to know if the cost in the two groups are the same.
- Since we have two independent samples, could use two-sample t-test
- Notice that the two graphs do not appear normal and have many outliers





Wilcoxon rank sum test

- Since we have two independent samples and the t-test is not appropriate, we need a nonparametric test. Unfortunately, statisticians are not too clever, so they named the test for two independent samples Wilcoxon rank sum.

- Again, we are interested in the median rather than the mean.

- The hypothesis test of interest is
  - H_0: $\text{median}_{\text{splenectomy}} = \text{median}_{\text{splenic preservation}}$
  - H_A: $\text{median}_{\text{splenectomy}} \neq \text{median}_{\text{splenic preservation}}$

Wilcoxon rank sum test

- Again, we use the rank of the data points, rather than the actual values.

- Under the null hypothesis, the number of high and low ranks in each group should be equal. If the sum of the ranks in one group is very high or very low, this would be evidence against the null hypothesis

Splenectomy	Rank	Splenic preservation	Rank
1011.07	10	496.44	4
1066.82	11	541.76	6
610.80	8	1562.01	19
1111.44	12	2515.12	21
955.68	9	1133.99	13
1203.84	14	300.33	2
1600.32	20	482.55	3
555.90	7	503.22	5
1302.95	17	2744.23	22
182.34	1	1232.22	15
1233.20	16		
1402.09	18		

Hypothesis test

- Two independent samples, Wilcoxon test, alpha=0.05

- Hypotheses
  - Null: $\text{median}_{\text{splenectomy}} = \text{median}_{\text{splenic preservation}}$
  - Alternative: $\text{median}_{\text{splenectomy}} \neq \text{median}_{\text{splenic preservation}}$

- Test statistic: Sum of positive ranks = 34,   p-value=0.77

- Fail to reject null hypothesis

- Conclusion: There is no evidence of a difference between the cost of anesthesia in the splenectomy patients and the splenic preservation patients.

Parametric tests – Nonparametric tests Equivalent

- Paired t-test – **Wilcoxon signed rank**

- Two sample t-test – **Wilcoxon rank sum (Mann-Whitney Test)**

- ANOVA – **Kruskal-Wallis** test

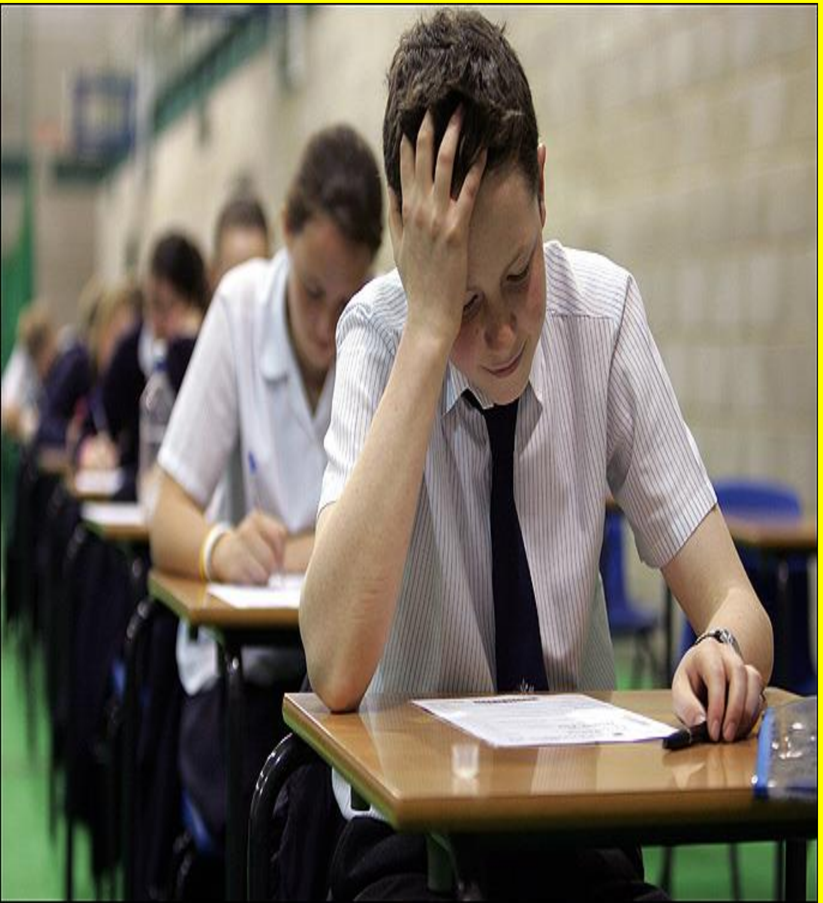
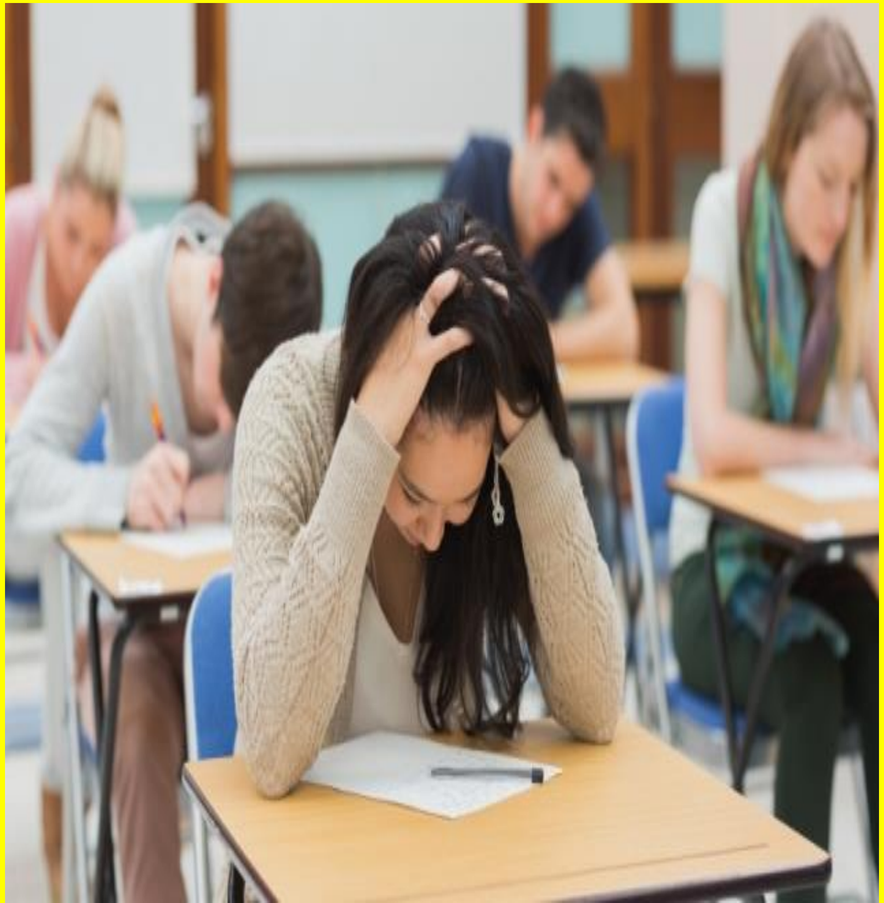
When you have two or more independent samples and the assumptions of ANOVA are not met, you can use the Kruskal-Wallis test. **This is a rank based test**.

Students were given different drug treatments before revising for their exams.

Some were given a memory drug, some a placebo drug and some no treatment

The exam scores (%) are shown below for the three different groups:




Carry out a one-way ANOVA by hand to test the hypothesis that the treatments will have different effects.

Can you suggest some follow-up comparisons?

Medical Treatment	Placebo	No Treatment
70	10	57
37	83	23
3	50	97
77	17	63
43	90	30


49
المدينة الجامعية ـ الشارقة
UNIVERSITY CITY - SHARJAH

Three different traffic routes are tested for mean driving time. The entries in the table are the driving times in minutes on the three different routes

University Road

Emirates Road

Airport Road

Which route is faster to the University?

Route 1	Route 2	Route 3
30	27	16
32	29	41
27	28	22
35	36	31

The following are the satisfaction scores of customers at the branches of a fast food restaurant

Is there a significant difference in satisfaction of the customers **among the 3 branches**?

Dine-in Restaurant Survey

1. In a typical month, how often do you visit our restaurant?

- ○ This is my first visit
- ○ Once or twice
- ○ More than four times

2. Rate your satisfaction with:

Taste of food	○	○	○	○	○
Quality of ingredients	○	○	○	○	○
Food service speed	○	○	○	○	○
Food temperature	○	○	○	○	○

3. Please rate your preferences:

Overall ambience	○	○	○	○	○
A/C	○	○	○	○	○
Music	○	○	○	○	○
Restaurant cleanliness	○	○	○	○	○
Restrooms cleanliness	○	○	○	○	○

4. Would you to recommend our restaurant to a friend or relative?

○ Yes ○ No

5. Do you have any suggestions to improve our quality of service?

Carrefour	Sahara Mall	Safeer Mall
1	3	1
2	2	3
3	3	2
3	3	1
2	2	3
4	3	1
2	2	1
4	3	3
2	2	1
3	1	2
2	1	3
3	1	2



The average of grade point averages (GPAs) of college courses in a specific major is a measure of difficulty of the major. An educator wishes to conduct a study to find out whether the difficulty levels of different majors are the same. For such a study, a random sample of major grade point averages (GPA) of 11 graduating seniors at a large university is selected for each of the four majors mathematics, English, education, and biology. The data are given in the Table.

Test, at the 5% level of significance, whether the data contain sufficient evidence to conclude that there are differences among the average major GPAs of these four majors.

Mathematics	English	Education	Biology
2.59	3.64	4.00	2.78
3.13	3.19	3.59	3.51
2.97	3.15	2.80	2.65
2.50	3.78	2.39	3.16
2.53	3.03	3.47	2.94
3.29	2.61	3.59	2.32
2.53	3.20	3.74	2.58
3.17	3.30	3.77	3.21
2.70	3.54	3.13	3.23
3.88	3.25	3.00	3.57
2.64	4.00	3.47	3.22