SELECTED NONPARAMETRIC TESTS

for inferential procedures: hypothesis ~~tests~~ and confidence intervals

Question: What $\boxed{\text{do we do}}$ when the underlying distributions of data are not normal?

(1) Ignore the problem, perhaps because the deviations from normality are not that great.

(2) If the "problem" is due to a few outliers, simply remove them.

(3) Sometimes, we can replace the assumption that the data come from a normal population, yet, we can still come up with theoretical results about the test statistic.

EX: $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

(4) There are "non-normality-based approaches" or "empirical approaches" — such as bootstrapping and permutation testing — which we have been getting a taste of in recent homework assignments.

(5) Finally, there are nonparametric methods that do not assume any specific form for the distribution of the population. They tend not to make use of the actual values of the observations but rather some function of those values (e.g., the ranks).

DEF: An inferential procedure is parametric if it depends on some assumption about the underlying distribution of the members of the simple random sample — i.e., a distribution that depends on a finite number of parameter.

EX: If X_i, $i=1,\dots,n$, has an exponential distribution with parameter λ, X_i iid.

EX: If $X_1,\dots,X_n$ is an i.i.d. simple random sample from a normal random variable.

$$X_i \sim N(\underbrace{\mu,\sigma^2}_{\text{2-parameter family}})$$

NOTE: If an inferential procedure is non-parametric, that means that there are very few (if any) assumptions about the underlying distribution of the data.

↳ Sometimes this means that you assume an infinite number of parameters.

THE WILCOXON RANK-SUM TEST

→ Used for comparing two samples, i.e., testing $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$, where $\tilde{\mu}_i$ is the median of the i^{th} population.

→ If these populations are assumed to have (presumably non-normal) symmetric distributions, then $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$ becomes a test of $H_0: \mu_1 = \mu_2$.

→ Let's use an example to explore the test.

EXAMPLE: Eight acres of land are seeded with corn. On four plots, no weeds are allowed. On the four plots, an average of three weeds per meter are allowed. The yields are measured at harvest time and are found to be:

0 weeds/meter	3 weeds/meter
166.7	158.6
172.2	176.4
165.0	153.1
176.9	156.0

Quickly eyeballing the data: it seems as if no weeds are associated to greater yields. But, we should be formal, and perhaps we are uncomfortable with an underlying assumption of normality.

Let's rank the aggregated sample:

153.1	156.0	158.6	165.0	166.7	172.2	176.4	176.9
1	2	3	4	5	6	7	8

0 weeds/meter group

Replace the quantitative data with ranks

NOTE: Replacing the quantitative data with ranks allows us to dispense with specific conditions related to the shapes of the underlying distributions.

NOTE: Under H_0, if no difference between weeds and no weeds exists, the mean (or the sum) of their respective ranks should be about equal.

NOTE: Sum of all the ranks from a data set (aggregated) of size n is always $\frac{n(n+1)}{2}$.

So....

Group	Sum of ranks
0 weeds/meter	23
3 weeds/meter	13

this will be our test statistic!

FORMAL STATEMENT OF THE WILCOXON RANK-SUM TEST

Draw a simple random sample of size n_1 from the first population and of size n_2 from the second population. Rank all $N = n_1 + n_2$ observations. Call the sum of the ranks from the first population by W. Under $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$, W is

1. approximately normal
2. has $\mathbb{E}[W] = \mu_W = \dfrac{n_1(n_1+n_2+1)}{2}$
3. has $\mathrm{Var}(W) = \sigma_W^2 = \dfrac{n_1 n_2 (n_1+n_2+1)}{12}$.

$$\frac{W - \mu_W}{\sigma_W} \sim N(0,1)$$

In the example involving the corn,

$$\mu_W = 18 = \frac{4(4+4+1)}{2}$$

$$\sigma_W = \sqrt{\frac{(4)(4)(9)}{12}} = \sqrt{12} \approx 3.464.$$

NOTE: $\mathbb{P}(W \geq 23) = \mathbb{P}\left(\dfrac{W-\mu_W}{\sigma_W} \geq \dfrac{23-18}{3.464}\right)$

$$= \mathbb{P}(Z \geq 1.4434) \approx 0.074 > \alpha = 0.05$$

$H_0: \tilde{\mu}_1 = \tilde{\mu}_2$

$H_1: \tilde{\mu}_1 > \tilde{\mu}_2$

$\Rightarrow$ and so we do not reject H_0.

At this time, we lack evidence to conclude that an absence of weeds improves corn yield.

NOTE: Some textbooks or treatments want to work with a continuity correction that makes W closer to normal r.v. for smaller values of n. Namely, we always agree to conservatize W by treating it as the smallest value that would round ~~up~~ to the sum of all the all the ranks. In our example,

$$\boxed{\cancel{\mathbb{P}\left(\frac{W-\mu_W}{\sigma_W} \geq \frac{23-18}{3.464}\right)}}$$

$$\mathbb{P}\left(\frac{W-\mu_W}{\sigma_W} \geq \frac{22.5-18}{3.464}\right)$$

As N becomes large, convergence to a normal r.v. is pretty fast, i.e., continuity correction is really only helpful when N is very small (e.g., 8).

NOTE: Some textbooks state H_0 differently...

$H_0: \tilde{\mu}_1 = \tilde{\mu}_2$ ← approach when ~~you~~ you can assume the populations have same shape except for the medians

probably more technically correct → H_0: two pop. distributions have the same shape

NOTE: What about ties in rankings?

⟹ Replace all tied values with the average of the ranks they occupy.

EX:

Observation	153	155	158	158	158	161	164	164	168
Ranks	1	2	4	4	4	6	7.5	7.5	9

WILCOXON SIGNED RANK TEST

Qn: What if we do not have independent samples — but rather, dependent samples — and we are not sure if the differences are normal?

[ASIDE: An analogue to the paired t test.]

EX: Each child is told two stories. The first story is read to them and the second is read as well but also illustrated with pictures. They are given scores on the recounting of each story. The data are as follows:

	Child 1	2	3	4	5
Story 2	0.77	0.49	0.66	0.28	0.38
Story 1	0.40	0.72	0.00	0.36	0.55
Differences	0.37	-0.23	0.66	-0.08	-0.17

Suppose that we wanted to test:

H_0: Scores have the same distribution for both stories

H_1: scores are symmetrically higher for story 2

[NOTE: Matched pairs t test gives $t = 0.635$ with one-sided p-value of 0.280.]

Approach is to rank the absolute values

Absolute Value	0.08	0.17	0.23	0.37	0.66
Rank	1	2	3	4	5

$\underbrace{1 \quad 2 \quad 3}_{\text{negatively signed}} \quad \underbrace{4 \quad 5}_{\text{positively signed}}$

The test statistic is the sum of the ranks of the positive differences and it tends to be called the Wilson Signed rank statistic.

It is sometimes denoted W^+.

Under H_0, W^+ is approximately normally distributed with mean

$$\mu_{W^+} = \frac{n(n+1)}{4}$$

total # of pairs we are working with

and standard deviation

$$\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

TESTING PHILOSOPHY: when W^+ is far above its mean, we reject H_0.

Going back to our specific case,

$$\mu_{W^+} = \frac{n(n+1)}{4} = \frac{5 \cdot 6}{4} = 7.5$$

$$\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{5 \cdot 6 \cdot 11}{24}} = \sqrt{13.75} \approx 3.708$$

$$\Rightarrow \mathbb{P}(W^+ \geq 9) = \mathbb{P}\left(\frac{W^+ - \mu_{W^+}}{\sigma_{W^+}} \geq \frac{9 - 7.5}{3.708}\right)$$

$$= \mathbb{P}(Z \geq 0.404) \approx 0.343 > \alpha = 0.05$$

$\Rightarrow$ Again, we do not reject H_0. There is not sufficient evidence at this time to conclude that pictures are associated with a difference in story recounting skills.

Qn: Does this give us a road map for testing $H_0: \mu = \mu_0$, i.e., the one-sample t test?

1. Take data and subtract $\tilde{\mu}_0$, the null-hypothesized median.
2. Sort the absolute values of the transformed data and rank it, breaking ties appropriately.
3. ~~Be sure to note whether or not the member of each absolute difference's ranking~~ Be sure to not the groups from which the absolute values of the data (minus $\tilde{\mu}_0$) come.
4. Calculate W^+ and proceed as before...

THE PEARSON CORRELATION COEFFICIENT

ISSUE: The standard (linear, Pearson) correlation coefficient is sensitive to departures from normality. This has consequences for ~~[illegible]~~ inferential matters.

THE ALTERNATIVE: Take the data — both independent $x_1, \dots, x_n$ and dependent data $y_1, \dots, y_n$ — and replace both sets of data with their ranks. Compute the sample correlation coefficient between those ranks!

in the usual Pearsonian sense

(It is called Spearman's rho and is sometimes denoted r or ρ, and sometimes has a hat.)

I will use $\hat{r}$ to denote the calculation of Spearman's rho for observed data.

NOTE: It can be shown that $Var(\hat{r}) \approx \dfrac{0.6325}{\sqrt{n-1}}$

this is the number of pairs.

NOTE: If $H_0: r = r_0$, $\dfrac{\hat{r} - r_0}{\sqrt{Var(\hat{r})}} \sim N(0,1)$.

KENDALL'S τ TAU

.... is another nonparametric alternative to Pearson's ρ.

DEFINITION: A pair of data (x_i, y_i) and (x_j, y_j) is said to be concordant if either

① $x_i < x_j$ and $y_i < y_j$

-or-

② $x_i > x_j$ and $y_i > y_j$

DEFINITION: A pair of data (x_i, y_i) and (x_j, y_j) is said to be discordant if it is not concordant.

DEFINITION: Kendall's tau is calculated as

$$\hat{\tau} = \frac{(\#\text{ of concordant pairs} - \#\text{ of discordant pairs})}{\binom{n}{2}}$$

NOTE: $\mathbb{E}[\hat{\tau}] = \tau$

$$\mathrm{Var}(\hat{\tau}) = \frac{2(2n+5)}{9n(n-1)}$$

and $\hat{\tau}$ is approximately normal, and so

$$\frac{\hat{\tau} - \tau_0}{\sqrt{\mathrm{Var}(\hat{\tau})}} \sim N(0,1)$$