Second Course in Statistics: lecture 18

Nonparametric tests

- Wilcoxon rank sum test
- Wilcoxon signed rank test

Inference for two way-way tables

- Analysis of two way tables
- Inference of two way tables
- Chi-square test of independence

Assumption of parametric test

The most commonly used methods for inference about the means of quantitative response variables assume that the variables in question have normal distributions in the population or populations from which we draw out data. In practice, of course, no distribution is exactly normal.

The one-sample and two-sample t procedures and analysis of variance are quite Robust. That is , the results of inference are not very sensitive to moderate lack of normality, especially when the samples are reasonably large.

Question

What can we do if graphic presentations suggest that the data are clearly not normal, especially when we have only a few observations.

Statistical inference

Comparison: parametric and nonparametric test

Setting	Normal setting	Rank test
One sample	one-sample t test	Wilcoxon signed rank test
Matched pair	one-sample t test to the differences	Wilcoxon signed rank test
Two independent samples	two-sample t test	Wilcoxon rank sum test
Several independent samples	one-way ANOVA F test	Kruskal-Wallis test

Statistical inference

Nonparametric test: Assumptions and hypothesis

Assumptions

Rank tests are designed to replace the t tests and one-way analysis of variance when normality conditions for those tests are not met. All of these tests require that the distribution of population or populations is continuous and can be described by a density curve.

Normal curves are one shape of density curve. Rank tests allow curves of any shape.

Hypothesis

The rank tests concern the center of a population or populations. When a population has at least roughly a normal distribution, we describe its center by mean. When the distributions are strongly skewed, we often prefer the median to the mean as a measure of center. In simplest form, the hypotheses for rank tests just replace mean by median.

Statistical inference

Nonparametric test: Rank transformation

Moving from the original observations to their ranks is a transformation of the data, the same as moving from the observations to their logarithms. The rank transformation retains only the ordering of the observations and makes no other use of their numerical values.
Working with ranks allows us to dispense with specific assumptions about the shape of the distribution, such as normality.

Statistical inference

Definition

- In a single sample, order the observations from the smallest to the largest. The rank of each observation is its position in this ordered list, starting with rank 1 for the smallest observation.
- For more than one samples, we need to rank observations in a combined sample. First arrange them in order from the smallest to the largest and then assign rank the same as for a single sample.

Definition

Draw an SRS of size n_1 from one population and draw an independent SRS of size n_2 from a second population. There are N observations in all, where $N = n_1 + n_2$. Rank all N observations. The sum W of the ranks for the first sample is the Wilcoxon rank sum statistic. If the two populations have the same continuous distribution (under H_0), then W has

$$\mu_W = \frac{n_1(N+1)}{2} \quad \text{and} \quad \sigma_W^2 = \frac{n_1 n_2 (N+1)}{12}$$

The Wilcoxon rank sum test rejects null hypothesis that the two populations have identical distributions when the rank sum W is far from its mean.

Example

Does the presence of small numbers of weeds reduce the yield of corn? Lamb's-quarter is a common weed in corn fields. A researcher planted corn at the same rate in 8 small plots of ground, then weeded the corn rows by hand to allow no weeds in 4 randomly selected plots and exactly 3 lamb's-quarter plants per meter of row in the other plots. Here are the yields of corn (bushels per acre) in each of the plots:

Weeds per meter	Yield (bu/acre)			
0	166.7	172.2	165.0	176.9
3	158.6	176.4	153.1	156.0

Solution

- Rank transformation:

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

- Sum of ranks:

	Sum of ranks
No weeds	23
3 weeds	13

Solution continued

- H_0: no difference in distribution of yields
  H_a: yields are systematically higher than weed-free plots
- T.S. : Let W be sum of rank for weed free plots, and

$$\mu_W = \frac{n_1(N+1)}{2} = \frac{4(8+1)}{2} = 18$$

$$\sigma_W = \sqrt{\frac{n_1 n_2 (N+1)}{12}} = \sqrt{\frac{4 \times 4 \times 9}{12}} = \sqrt{12} \approx 3.464$$

- R.R. : If W is 2 σ's away from μ_W, reject H_0.

Solution continued

- Conclusion:
  Since $\frac{W-\mu_X}{\sigma} = \frac{23-18}{3.464} \approx 1.44$ i.e. observed rank sum w is only about 1.4 standard deviation higher than μ_W, we now suspect that the data do not give strong evidence that yields are higher in the population of weed-free corn.

The normal approximation

The rank sum statistic W becomes approximately normal as the two sample sizes increase. We can then form yet another Z statistic by standardizing W:

$$Z = \frac{W - \mu_W}{\sigma_W} = \frac{W - n_1(N+1)/2}{\sqrt{n_1 n_2 (N+1)/12}}$$

Continuity correction:

Continuity correction improves the accuracy of the approximation. E.g.

$$P(W \geq 23) = P(W \geq 22.5) \qquad P(W \leq 13) = P(W \leq 13.5)$$

Handling ties

The exact distribution for the Wilcoxon rank sum is obtained assuming that all observations in both samples take different values. This allows us to rank them all. In practice, however, we often find observations tied at the same value. The usual practice is to assign all tied values the average ranks they occupy. Here is an example with 6 observations:

Obervation	153	155	158	158	161	164
Rank	1	2	3.5	3.5	5	6

The exact distribution for the Wilcoxon rank sum W changes if the data contain ties. Moreover, the standard deviation σ_W must be adjusted if ties are present.

Statistical inference

Nonparametric test: Wilcoxon signed rank test

Review:

We use the one-sample t procedures for inference about the mean of one population or for inference about the mean difference in a matched pairs setting. The matched pairs setting is more important because good studies are generally comparative.

Paired t test requires that the difference is normally distributed. When the differeces are not normal or we lack of the information of normality assumption from the sample values, Wilcoxon signed rank test is designed to cope with such an issue.

Definition

Draw an SRS of size n from a population for a matched pairs study and take the differences in responses within pairs. Rank the absolute values of these differences. The sum W^+ of the ranks for the positive differences is the Wilcoxon signed rank statistic. If the distribution of the responses is not affected by the different treatments within pairs, then W^+ has

$$\mu_{W^+} = \frac{n(n+1)}{4} \quad \text{and} \quad \sigma^2_{W^+} = \frac{n(n+1)(2n+1)}{24}$$

The Wilcoxon signed rank test rejects the null hypothesis that there are no systematic differences within pairs when the rank sum W^+ is far from its mean.

Statistical inference

Nonparametric test: Wilcoxon signed rank test

Example

A study of early childhood education asked kindergarten students to retell two fairy tales that had been read to them earlier in the week. Each child told two stories. The first had been read to them, and the second had been read but also illustrated with pictures. An expert listened to a recording of the children and assigned a score for certain uses of language. Here are the data for five ”low-progress” readers in a pilot study:

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
Difference	0.37	-0.23	0.66	-0.08	-0.17

We wonder if story with illustrations improves how the children retell a story.

Solution

- Rank of absolute differences:

absolute difference	0.37	0.23	0.66	0.08	0.17
Rank	4	3	5	1	2

- H_0: There is no difference between scores for both stories.
  H_a: Scores are systematically higher for story 2.
- T.S. : Let W^+ be sum ranks of positive difference.
  $W^+ = 4 + 5 = 9$
  $$\mu_{W^+} = \frac{n(n+1)}{4} = \frac{5 \times 6}{4} = 7.5$$
  $$\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{5 \times 6 \times 11}{24}} \approx 3.708$$

Statistical inference

Nonparametric test: Wilcoxon signed rank test

Solution continued

- R.R.
  - Method 1: Reject H_0 if observed w^+ is two standard deviations away from the mean
  - Method 2: Assume W^+ normally distributed with μ_{W^+} and σ_{W^+}, then calculate p-value.

$$\begin{aligned}\text{P-value} &= P(W^+ \geq 9) \Longrightarrow P(W^+ \geq 8.5) \\ &= P\left(Z \geq \frac{8.5 - 7.5}{3.708}\right) = P(Z > 0.27) \approx 0.394\end{aligned}$$

- Conclusion: Since p-value $> 5\%$, we cannot reject H_0 that the score distributions are the same for both stories.

The normal approximation:

Wilcoxon signed rank test statistic W^+ becomes approximately normal as the two sample sizes increase. We can then form yet another Z statistic by standardizing W^+:

$$Z = \frac{W^+ - \mu_{W^+}}{\sigma_{W^+}} = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

Continuity correction is also needed to improve accuracy.

Handling ties

- Ties among the absolute differences are handled by assigning average ranks.
- A tie within a pair creates a difference of zero. Because these are neither positive nor negative, the usual procedure simply drops such pairs from the sample. This amounts to dropping observations that favor the null hypotheis (no difference). If there are many ties, the test may be biased in favor of the alternative hypothesis.
- As in the case of the Wilcoxon rank sum, ties distort the distribution of W^+ and complicate finding a p-value. The standard deviation σ_{W^+} must be adjusted for the ties before we can use the normal approximation. Software will do this.

Review:

From previous lectures inference about proportions in one-sample and two-sample settings deals with the binary outcomes in sequence of n independent Bernoulli distributions.

We now study how to compare two or more populations when response variable has two or more possible values and how to test whether two categorical variables are independent. We need to begin our analysis with two-way tables.

Example

Two-way tabulation of 200 employees by employee classification and opinion on collective bargaining (Ott and Mendenhall 6th, ed. p412)

Employee classification	Opinion on collective bargaining: Favor	Not favor	Undecided	Total
Staff	30	15	15	60
Faculty	40	50	10	100
Administrator	10	25	5	40
Total	80	90	30	200

Employment example continued

Row percentage Employee classification	Opinion on collective bargaining Favor	Not favor	Undecided	Total
Staff	50%	25%	25%	100%
Faculty	40%	50%	10%	100%
Administrator	25%	62.5%	12.5%	100%

Employment example continued

Column percentage	Opinion on collective bargaining		
Employee classification	Favor	Not favor	Undecided
Staff	37.5%	16.67%	50%
Faculty	50%	55.55%	33.33%
Administrator	12.5%	27.78%	16.67%
Total	100%	100%	100%

Statistical inference

Analysis of two-way table

Concept of independence:

Two variables that have been categorized in a two-way table are independent if the probability that a measurement is classified into a given cell of the table is equal to the probability of its being classified into that row times the probability of its being into that column

This must be true for all cells of the table.

Employment example continued

If employee classification and opinion are independent, the cell probabilities are

Employee classification	Opinion on collective bargaining Favor	Not favor	Undecided	Total
Staff	$p_A p_1$	$p_A p_2$	$p_A p_3$	p_A
Faculty	$p_B p_1$	$p_B p_2$	$p_B p_3$	p_B
Administrator	$p_C p_1$	$p_C p_2$	$p_C p_3$	p_C
Total	p_1	p_2	p_3	1

In practice we do not use probability directly and use frequency instead.

Observed and marginal frequencies

	A_1	A_2	$\cdots$	A_k	$\sum$
B_1	O_{11}	O_{21}	$\cdots$	O_{k1}	$n_{.1}$
B_2	O_{12}	O_{22}	$\cdots$	O_{k2}	$n_{.2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_r	O_{1r}	O_{2r}	$\cdots$	O_{kr}	$n_{.r}$
$\sum$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{k.}$	$\mathbf{n}$

O_{ij} is the observed frequency.

$n_{i.}$ and $n_{.j}$ are the marginal frequency

Statistical inference

Analysis of two-way table

Expected frequencies

	A_1	A_2	$\cdots$	A_k	$\sum$
B_1	E_{11}	E_{21}	$\cdots$	E_{k1}	$n_{.1}$
B_2	E_{12}	E_{22}	$\cdots$	E_{k2}	$n_{.2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_r	E_{1r}	E_{2r}	$\cdots$	E_{kr}	$n_{.r}$
$\sum$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{k.}$	**n**

Where $E_{ij} = np_{ij} = np_{i.}p_{.j} = n \times \frac{n_{i.}}{n} \times \frac{n_{.j}}{n} = \frac{n_{i.} \cdot n_{.j}}{n}$

Employment example continued: expected cell frequencies in blue

Employee classification	Opinion on collective bargaining: Favor	Not favor	Undecided	Total
Staff	30 (24)	15 (27)	15 (9)	60
Faculty	40 (40)	50 (45)	10 (15)	100
Administrator	10 (16)	25 (18)	5 (6)	40
Total	80	90	30	200

Where $E_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{80 \times 60}{200} = 24$

Statistical inference

Chi-square test of independence

Assumptions

Two random variables are the measurements on the data from one single population.

- E_{ij} for each class must be larger than 1
- No more than $20\% E'_{ij}s$ are less than 5.

Hypothesis

H_0 : X and Y are independent, or H_0: $p_{ij} = p_{i\cdot}p_{\cdot j}$ for all pairs i, j
H_a: X and Y are dependent. or H_a: $p_{ij} \neq p_{i\cdot}p_{\cdot j}$ for some pairs i, j

Test statistic:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \text{or} \quad \chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{r} \frac{O_{ij}^2}{E_{ij}} - n$$

Degree of freedom $\nu = (k-1)(r-1)$

Rejection region:

Reject H_0 if $\chi^2 > \chi^2_\alpha[(k-1)(r-1)]$ at α level of significance.

Employment example continued

Test the independence of employee classification and opinion

Solution

H_0: Employment status is independent of opinions.
H_a: Employment status s NOT independent of opinions.

Statistical inference

Chi-square test of independence

Solution continued

T.S.:

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \frac{(30-24)^2}{24} + \frac{(15-27)^2}{27} + \frac{(15-9)^2}{9} + \frac{(40-40)^2}{40}$$

$$+ \frac{(50-45)^2}{45} + \frac{(10-15)^2}{15} + \frac{(10-16)^2}{16} + \frac{(25-18)^2}{18}$$

$$+ \frac{(5-6)^2}{6} = 18.2$$

$$\nu = (k-1)(r-1) = (3-1)(3-1) = 4$$

Solution continued

- R.R.: Reject H_0 if $\chi^2 > \chi^2_{.05}(4) = 9.488$
- Conclusion: Since observed $\chi^2 = 18.2 > 9.488$, we reject H_0 at 5% level of significance and conclude that two variables are dependent.