Parametric Statistics

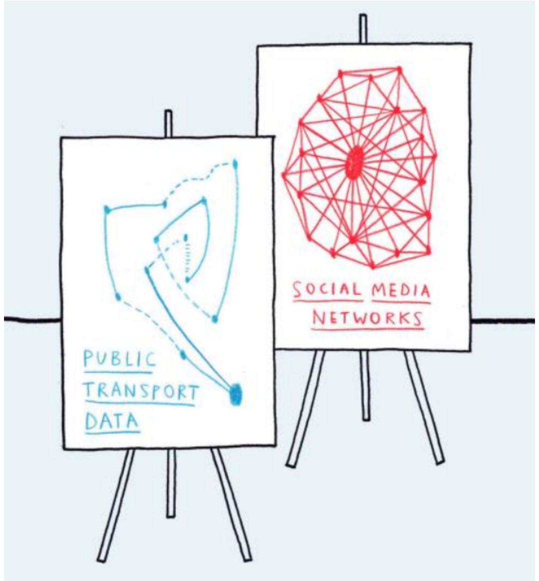
Week 6 - Non-parametric Tests

Juan Carlos Medina Serrano

Technische Universität München
Hochschule für Politik
Political Data Science

Munich, 27. November 2019

political data science
https://politicaldatascience.blogspot.de



Non-normality

What happens when the assumption of normality does not hold?

- **Transform the data** (make it normal)
- **Non-parametric tests**
- **Resampling Procedures** (Bootstraping, Permutation tests)

Transforming the Data

A transformation to make the data normal won't change the relationships between variables, but it does change the differences between different variables.

It is important to transform all variables in the analysis (for example, when comparing means)

Transformations that can correct for **positive** skew and unequal variances:

- **Log transformation:** $\log(X)$.
- **Square root transformation:** $\sqrt{X}$.
- **Reciprocal transformation:** $1/X$. It reverses the value, large values become small and viceversa.

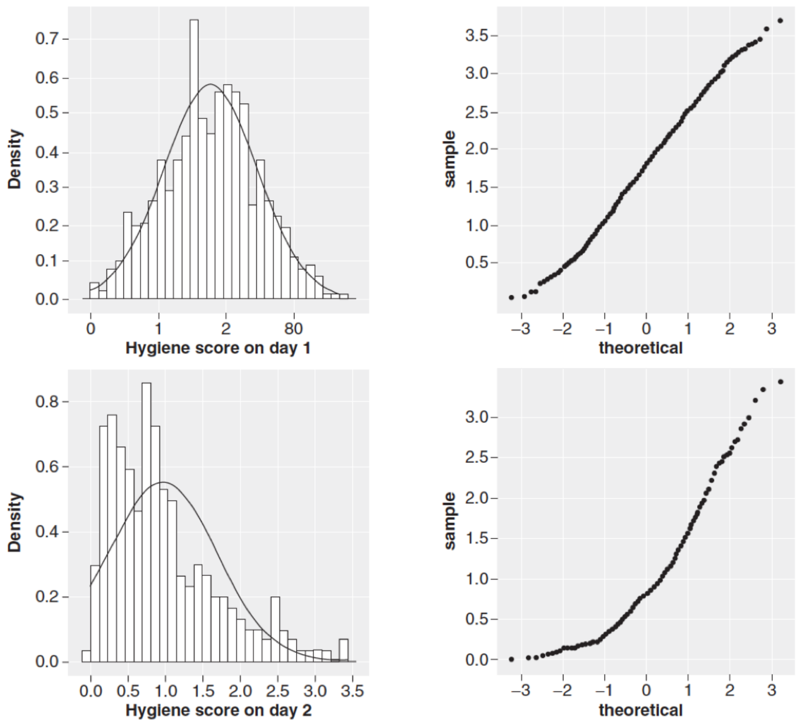
Transformations that can correct for **negative** skew and unequal variances:

- All of the above, but before, we need to reverse the scores: Maximum value minus the value: $X_{Highest} - X$

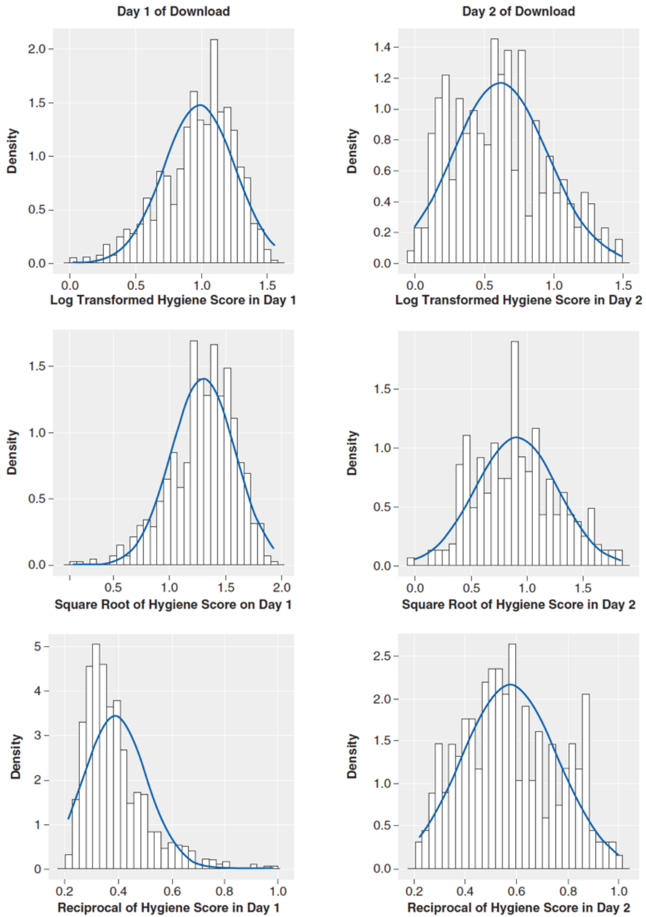
Transforming the data has limitations, use it with caution!

Transforming the Data

Example Remember the q-q plot?



Non-parametric Tests



Non-parametric Tests

Also known as assumption-free tests because they make fewer assumptions.

They mostly work on the principle of **ranking** the data:

Lowest value has the rank 1, the next value has rank 2 and so on.

If two observations have the same value, they become tied rank. Example: Rank 3 and 4 with the same value would become a rank 3.5

The statistical tests are then done on the rankings, not on the data itself.

Wilcoxon rank-sum test

Equivalent to the **independent t-test** : Compare means between two independent groups (control and treatment)

Idea: If we rank the data of the two groups together, what should we find?

- If there is no difference, there is a similar number of low and high ranks in each group.
- If there is a difference, one group will have more high ranks

If we added the ranks:

- If there is no difference, the summed total of ranks similar between groups
- If there is a difference, then one of the groups has a higher sum of ranks

Wilcoxon rank-sum test

Wilcoxon rank-sum test steps

1. Combine the values of the two groups
2. Rank the values
3. Group the ranks by group and sum them, select the lowest sum, we call it W.
4. Calculate the **mean rank** $\overline{W}$: This is simply the sum of the ranks of the first half of the ranks (for two groups of five, the mean rank is $1+2+3+4+5$) (for unequal group sizes a formula is used)
5. Calculate the **rank standard error** SE_W (Formula more complicated)
6. Perform a z-test!

$$z = \frac{W - \overline{W}}{SE_W}$$

In some way, we are comparing the **medians** of the two groups instead of means! Remember that the median is robust since its unaffected by extreme values

Wilcoxon signed-rank test

Equivalent to the **dependent t-test** : Compare means between two dependent groups (same participants)

Idea: We rank the differences between the after group and the before group.
We then sum the ranks, BUT the ranks of positive and negative differences separately.

Wilcoxon signed-rank test

Wilcoxon signed-rank test steps

1. Calculate the difference between after and before group.
2. Take the **absolute value** of the distances and rank them
3. Group the ranks by sign (positive and negative) and sum them. Take lowest value. We call it T.
4. Calculate the **mean rank** $\overline{T}$: This is simply the sum of ranks of the first half of the ranks divided by two (for a group of five subject tested before and after, the mean rank is $(1+2+3+4+5)/2$). We divide by two since they are the same participants.
5. Calculate the **rank standard error** SE_T (Formula more complicated)
6. Perform a z-test!

$$z = \frac{T - \overline{T}}{SE_T}$$

Wilcoxon signed-rank test

Example Alcohol test on Sunday and Wednesday for a group of 10 students.

BDI Sunday	BDI Wednesday	Difference	Sign	Rank	Positive ranks	Negative ranks
Alcohol						
16	5	−11	−	9		9
15	6	−9	−	7		7
20	30	10	+	8	8	
15	8	−7	−	3.5		3.5
16	9	−7	−	3.5		3.5
13	7	−6	−	2		2
14	6	−8	−	5.5		5.5
19	17	−2	−	1		1
18	3	−15	−	10		10
18	10	−8	−	5.5		5.5
Total =					8	47

$$\overline{T} = (1+2+3+4+5+6+7+8+9+10)/2 = 27.5$$

$$z = \frac{8-27.5}{9.81} = -1.99$$

Absolute value bigger than 1.96 (0.05 significance level) => Reject the null that there was no effect!

Kruskal-Wallis test

Equivalent to **one-way ANOVA** : Compare means between more than two independent groups.

Same idea as the Wilcoxon rank-sum test, but with a different test statistic.

Kruskal-Wallis test

Kruskal-Wallis test steps

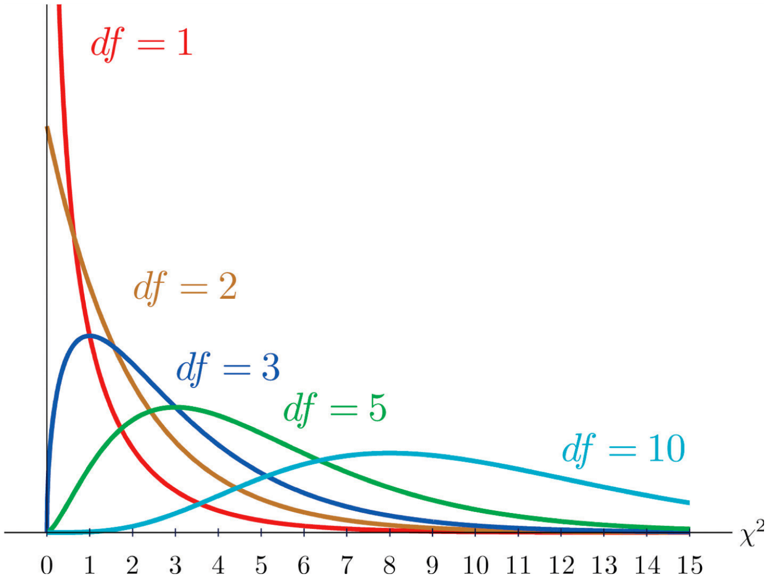
1. Combine the values of all the groups
2. Rank the values
3. Group the ranks by group and sum them.
4. Calculate the **H** statistic:

$$H = \frac{12}{N(N+1)} * \sum_{i}^{k} \frac{R_i^2}{n_i} - 3 * (N+1)$$

   With N total observations, and n_i the number of observations in group i. (Formula is shown here only to show you that test statistics can be really complicated, R does the hard work for us)
5. Compare with the critical value of the **chi-square** distribution, with $k-1$ degrees of freedom (k is the number of groups)

Chi-squared Distribution

Chi is a greek letter: χ
Has one parameter: k which specifies the number of degrees of freedom.



With very large k, the chi-square distribution approaches a normal distribution

Post-hoc Tests

As with one-way ANOVA, we need to do post-hoc tests to compare between each pair of groups.

We use the same correction as with ANOVA, the only difference is that we use Wilcoxon-rank sum tests between groups instead of independent t-tests.

Non-parametric Correlation

Pearson's correlation coefficient can only be used for normally distributed variables.

What can we do?
Wait for it...

Rank the data first!
Then calculate the Pearson correlation with the ranks!

Spearman's correlation coefficient!

$$\rho = \frac{cov(X_{Rank}, Y_{Rank})}{s_{Xrank} * s_{Yrank}}$$

The significance test is the same as with the Pearson correlation coefficient, where we used a t-test

Tests for Categorical Data

We finished with the non-parametric version of the statistical test we learned the last weeks!

Before we finish this chapter, lets look at two non-parametric tests that have no parametric equivalent:

- Chi-square test of goodness of fit
- Chi-square test of independence

They are tests for categorical variables with counts, and they both use the chi-square distribution to calculate significance.

Categorical variables have responses which consist of a set of categories.

Goodness of fit

This is a **ONE** variable test, that asks if the distribution of the categorical variable follows a specific pattern.

We can compare patterns with the proportion of the variables:

$$proportion = \frac{counts}{total\ counts}$$

One pattern is that all categories should have the same proportion (P). In the case of three categories:

$$H_0 : P_1 = P_2 = P_3$$

$$H_1 : P_1 \neq P_2 \neq P_3$$

Goodness of fit

We use this test statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where

- O_i is the observed counts in category i
- E is the expected counts in category i
- k is the number of categories

And we use the χ^2 distribution with k-1 degrees of freedom to calculate the critical value.

Goodness of fit

Example We interview 90 people on their Brexit preference, we want to test if the proportion of the three categories is the same.

For	Maybe	Against	Total
23	17	50	90

Goodness of fit

Example We interview 90 people on their Brexit preference, we want to test if the proportion of the three categories is the same.

For	Maybe	Against	Total
23	17	50	90

Category	O (observed frequency)	E (expected frequency)	D (difference)	$(O-E)^2$	$(O-E)^2/E$
For	23	30	7	49	1.63
Maybe	17	30	13	169	5.63
Against	50	30	20	400	13.33
Total	90	90			

Goodness of fit

Example We interview 90 people on their Brexit preference, we want to test if the proportion of the three categories is the same.

For	Maybe	Against	Total
23	17	50	90

Category	O (observed frequency)	E (expected frequency)	D (difference)	$(O-E)^2$	$(O-E)^2/E$
For	23	30	7	49	1.63
Maybe	17	30	13	169	5.63
Against	50	30	20	400	13.33
Total	90	90			

$$\chi^2 = 1.63 + 5.63 + 13.33 = 20.6$$

Critical value $= \chi^2$ distribution with $3 - 1 = 2$ degrees of freedom $= 5.99$

Reject the null that the proportions are the same

Goodness of fit

The null hypothesis does not have to be that the proportions are equal! We can design an experiment with specific proportions per class.

Example Blood pressure levels on 100 workaholics:

	Hypotension	Normal	Prehypertension	Hypertension	Total
Expected proportion	0.10	0.40	0.30	0.20	1.00
Expected count	10	40	30	20	100
Observed count	12	25	50	13	100

Chi-square test of independence

This is a **TWO** variable test, that asks if two categorical variables are independent.

Independent = No relation! (Equivalent to a correlation test between categorical values)

$$H_0 : Variables\ are\ independent = No\ relation$$

$$H_1 : Relation$$

We are in luck, we use the same statistic χ^2 as for the goodness of fit!

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

The only difference is how we calculate the expected count and the degrees of freedom.

Since we have two variables, degrees of freedom is number of categories - 1 in first variable multiplied by number of categories -1 in second variable.

Chi-square test of independence

We first put the data in a **contingency matrix** (one variable on the rows and one variable on the columns).

Example Is there a relationship between gender and voter participation on Brexit? (survey of 120 british people)

		Voter Participation	
		Vote	Don't Vote
Gender	Men	37	32
	Women	20	31

Chi-square test of independence

We calculate the expected counts as the multiplication of the row total by the column total divided by the total sum of all observations

	Men	Women	Total
Vote	37	32	69
Don't Vote	20	31	51
Total	57	63	120

The expected counts for Men who Vote is:

$$\frac{69 * 57}{120} = 32.77$$

Chi-square test of independence

Observed (O)			
	Men	**Women**	**Total**
Vote	37	32	69
Don't Vote	20	31	51
Total	57	63	120
Expected (E)			
	Men	**Women**	
Vote	32.78	36.23	
Don't Vote	24.23	26.78	
$(O - E)^2/E$			
	Men	**Women**	
Vote	0.54	0.49	
Don't Vote	0.74	0.67	

Chi-square test of independence

$(O-E)^2/E$			
	Men	**Women**	
Vote	0.54	0.49	
Don't Vote	0.74	0.67	

$$\chi^2 = .54 + .49 + .74 + .67 = 2.44$$

χ^2 distribution with $(2-1)*(2-1) = 1$ degrees of freedom, Critical value = 3.84

Unable to reject the null that gender and voter participation are independent = We can't prove that there is a relation between gender and voter participation

Resampling test of independence

Alternative data approach?

	Men	Women	Total
Vote	37	32	69
Don't Vote	20	31	51
Total	57	63	120

Resampling test of independence algorithm

1. Calculate $\sum_i (O_i - E_i)^2$ as before from the contingency table. This is the observed sum of squared differences.
2. Create a box with 57 ones (Men) and 63 twos (Women).
3. Shuffle the data, and take two samples **without** replacement of sizes 69 (Vote) and 51 (Don't vote)
4. Count the number of ones and twos in each sample
5. Fill this information in the contingency table and calculate $\sum_i (O_i - E_i)^2$
6. Repeat steps 2 to 5 R times (10000).
7. how often the resampled sum exceeds the observed sum of squared differences = p-value

Literature

- Discovering Statistics using R: Chapter 15
- Statistics for people who hate statistics: Chapters 17
- Parametrische Statistik: Chapter 5