Data analysis: Testing for differences

EE1622 Data and Information
Week 6
ECE, Brunel University London

Outline of lecture

- Choosing the right test for the data
- Difference tests
- T-test
  - Independent samples
  - Paired samples
- Non-parametric statistics

Relationships and Differences

- Researchers often want to know if there is a significant **relationship** between two variables
  - *Is there a relationship between* ***number of cats*** *one owns and* ***number of friends*** *one has?*
  - *Is the* ***amount spent playing computer games*** *related to* ***spatial skills*** *development?*
  - *Is there a relationship between* ***attendance rate*** *and* ***exam score****?*
  - *Does* ***gender*** *predict* ***early technology adoption****?*
- Or if there is a **difference** between one or more scores/conditions/groups
  - *Is* ***prototype A*** *more usable than* ***prototype B****?*
  - *Does* ***programming in pairs*** *improve performance?*
  - *Has the* ***new*** *system increased productivity?*
  - *Does* ***gender*** *affect the frequency and size of online*

I found a difference!

- Does **marital status** affect **life satisfaction**?
  - As measured through a questionnaire
- Does **gender** affect **memory?**
  - As measured by a memory test

Report

life sat

marital status	Mean	N	Std. Deviation
single	5.58	26	2.318
married/defacto	5.19	86	2.384
divorced	5.38	8	2.200
widowed	7.67	3	3.215
Total	5.34	123	2.381

Report

memory

sex	Mean	N	Std. Deviation
female	5.71	70	2.676
male	3.96	52	2.231
Total	4.97	122	2.635

- Great, but is the difference reliable/meaningful?
- How confident are you? Is it an accident (due to chance)?
- We need to have a statistical test!

Step 1: What is your question?

- Remember, when conducting research it is important to be clear about the questions you are trying to answer
  - Ideally before you begin data collection
- The questions…
  - *Is there a relationship between **attendance rate** and **exam score**?*
  - *Does **gender** predict **online purchase frequency?***
- …require quite different statistical tests to questions like:
  - *Is **prototype A** more usable than **prototype B**?*
  - *Does **group A** make more errors than **group B**?*
  - *Does **programming in pairs** improve performance?*

Tests of relationship vs. tests of difference

- Questions of the first kind are about relationships. Common techniques include:
  - **Correlation** – are variable X and variable Y related?
    - E.g. Head size and IQ, salary and job satisfaction
  - **Multiple regression** – which combination of independent variables offers the most accurate prediction of a given dependent variable?
    - E.g. A function of age and educational attainment may predict salary more accurately than either variable alone
  - **Chi-square** shows whether there is an association between two categorical variables (e.g. Is gender associated with smoking?)
- Questions of the second kind are about differences between groups/categories:
  - **T-tests** measure the difference between **two** groups (e.g. males, females) according to some *continuous variable* (e.g. Height, sales, job satisfaction)
  - **Analysis of variance** (ANOVA) measure differences when there *are **more than two groups** and/or more than two independent variables (e.g. marital status as in slide 4)*

- Last week we looked at **correlations**

I found a difference!

1. Does **marital status** affect **life satisfaction**?
   - As measured through a questionnaire
2. Does **gender** affect **memory?**
   - As measured by a memory test

Report

life sat

marital status	Mean	N	Std. Deviation
single	5.58	26	2.318
married/defacto	5.19	86	2.384
divorced	5.38	8	2.200
widowed	7.67	3	3.215
Total	5.34	123	2.381

Report

memory

sex	Mean	N	Std. Deviation
female	5.71	70	2.676
male	3.96	52	2.231
Total	4.97	122	2.635

- So what test to use for research question 1?
- What test to use for research question 2?

Step 2: Select your data

- Which variables will you be using?
- Which is the **independent variable (IV)**?
  - The variable that is believed to affect the dependent variable
  - What you control/manipulate
- Which is the **dependent variable (DV)**?
  - The observation that is believed to be affected by the IV
  - What you measure (aka outcome variable)
- Which is the IV and DV in the following questions:
  - Does **gender** affect **product ratings?**
  - Does **revision time** affect **test scores?**
  - Does the **website background colour** influence **reading speed**?
  - Which **type of interface** results in **higher user satisfaction**?

Step 2: Select your data

- What is the level of measurement for each variable?
  - Categorical (nominal, ordinal) or continuous (interval, ratio)?
  - Level of dependent measure will determine whether to apply a **parametric** or **non-parametri**c test variant
  - Level of independent measure may need to be reduced for difference tests (e.g. Collapsing *age* in years into a set of interval groups: 18-35, 36-49, 50+)

I found a difference

1. Does **marital status** affect **life satisfaction**?
   - As measured through a questionnaire
2. Does **gender** affect **memory?**
   - As measured by a memory test

Report

life sat

marital status	Mean	N	Std. Deviation
single	5.58	26	2.318
married/defacto	5.19	86	2.384
divorced	5.38	8	2.200
widowed	7.67	3	3.215
Total	5.34	123	2.381

Report

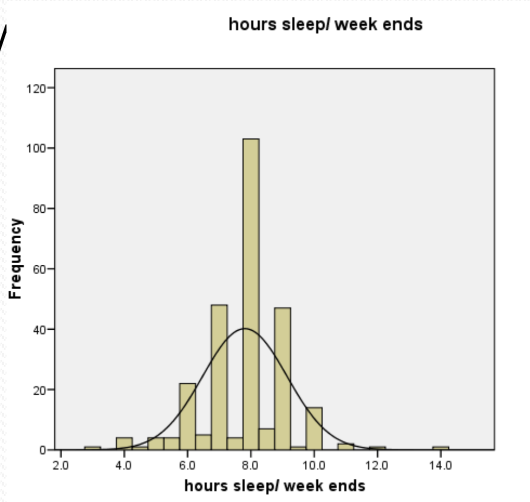
memory

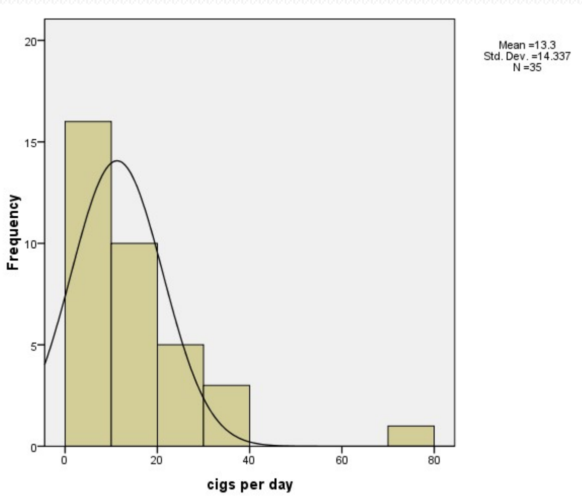
sex	Mean	N	Std. Deviation
female	5.71	70	2.676
male	3.96	52	2.231
Total	4.97	122	2.635

- So what are the DV and IV for research question 1?
- What is the level of measurement for each?
- What are the DV and IV for research question 2?
- What is the level of measurement for each?

Step 3: Know your data

- Descriptive statistics should be used to define the characteristics of your data (last week's lecture)
- For nominal data you need to know if numbers in each group/category are balanced
  - (e.g. Reliable comparison of gender effect not possible if 25 males and only 3 females)
- For continuous data you need to know if the distribution is normally distributed (e.g. Not skev





Step 4: Parametric vs. non-parametric

- **Parametric tests** make certain assumptions about the parameters (properties) of the sampled populations
  - Distributions are normal
  - Variances (square of standard deviation) of comparison groups are equivalent
  - Measurements are on a true continuous scale (equal interval or ratio level data)
- Most parametric tests are fairly robust i.e. tolerant of minor violations of assumptions, particularly if the sample size is quite large (i.e. N>100)
- However, if in doubt, use a non-parametric equivalent

- **Non-parametric tests** make no such assumptions about distribution
  - Analyse rank-order (ordinal properties) rather than continuous distribution
  - But are less powerful or more conservative (less likely to reject null hypothesis)

Difference tests

- From the table we can observe that *females appear to have better memory than males*
- The difference is quite small so
  - is it **significant** i.e. reliable?
  - What is the **effect size**?
- Significance and effect size depend not only on the size of the difference but also on:
  - The variability in the data – a small difference in means may be highly significant if the standard deviation is relatively small
  - The size of the sample – the larger the sample, the more reliable the estimates of the mean and standard deviation become

Report

memory

sex	Mean	N	Std. Deviation
female	5.71	70	2.676
male	3.96	52	2.231
Total	4.97	122	2.635

T-tests

- If we are testing the significance of an observed difference between two groups or treatment conditions we use the **t-test**
- **Independent-samples t-test** is used when comparing two different groups
  - You have a continuous (scale) level DV and a categorical (nominal/ordinal) IV
  - E.g. The example on the last slide: gender (IV) and memory score (DV)
- **Paired-samples t-test** is used when comparing the same subject (e.g. Person) on two different treatment conditions
  - e.g. Before and after some intervention: Memory before and after coffee consumption, Levels of stress before and after exercise,
  - Similar to correlation, you are comparing two scores from the same individual
  - However you wouldn’t do a paired t-test between age and height. **Why?**

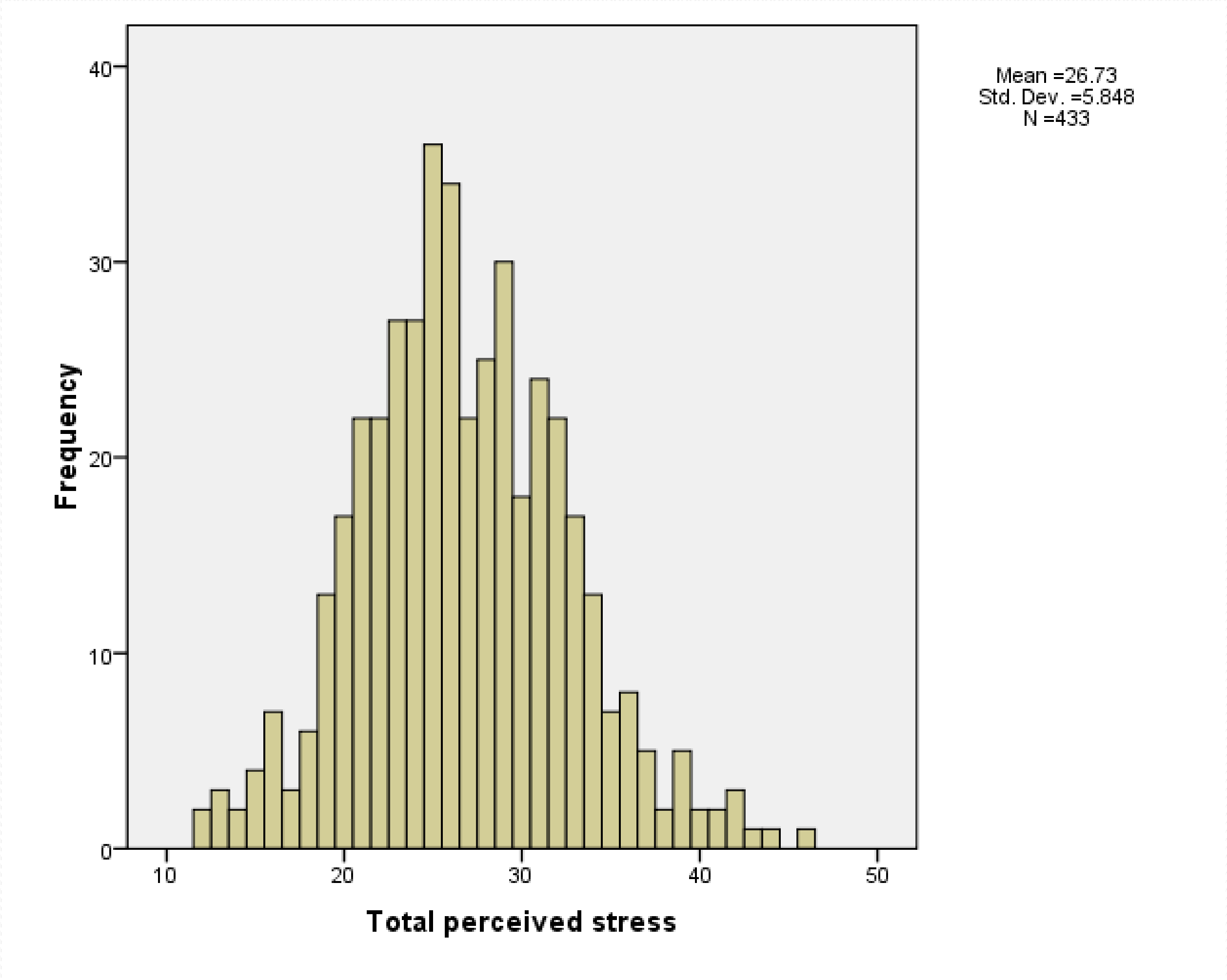
Assumptions to meet for parametric tests

- T-tests are parametric statistics so check for
  - Normality (skewness, kurtosis)
  - Level of measurement – continuous data only (usually)
  - Homogeneity of variance – i.e. Variance (SD) of groups being compared are equivalent
  - The data is a random sample of the population
  - Independence of observations – behaviour of individual subjects was not influenced by another subject in the sample. This can be a problem when data is collected in a group setting

SPSS: Independent t-test

- Question: *"Is there a significant difference in perceived* ***stress*** *for* ***males and females****?"*
  - *From Survey.sav*
- Go to: **Analyze → Compare Means → Independent Samples T-test**
- Move the DV into the "test variable" box
- Move the IV into the "grouping variable" box
- Click "Define groups" and type the code number for the first group into the "Group 1" box. In this case "Males = 1" so type "1". Enter the code for females into the "Group 2" box
- Click "Continue", followed by "OK"

Normality assumption met?



Output

- The “Group Statistics” table summarises the two groups:
- From the means, it appears that **males are less stressed than females.**
- The bottom table is more complicated
  - You need Levene’s test to be non-significant in order to meet the assumption of “Equal variances”
  - If the result of Levene’s test is significant (p <=0.05), then groups do not have equal variance, and you will need to use the second row of results “Equal variances not assumed”
  - In this example, as p = 0.165, this is not significant, so use the first row
- So **t = 2.898** and **p = 0.004** **(<0.05)**, with a mean difference of 1.634between groups
- **df** refers to **“degrees of freedom”**, which is based on N and used to determine significance
  - N1 + N2 – 2 in this case

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
Total perceived stress	FEMALES	249	27.42	6.066	.384
	MALES	184	25.79	5.414	.399

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Total perceived stress	Equal variances assumed	1.936	.165	2.898	431	.004	1.634	.564	.526	2.742
	Equal variances not assumed			2.948	415.886	.003	1.634	.554	.544	2.723

Interpreting the result

- p= 0.004
  - **there IS a statistically significant difference!!**
- Statistically significant difference is not enough.
  - You have to report the magnitude of this difference (the effect size)
- Checking the **effect size**
  - Eta squared measures how much of the variance in DV is explained by the IV
  - Eta squared = $t^2 / (t^2 + (N1+N2 – 2))$
  - $2.9^2 / (2.9^2 + (249 + 184 – 2))$ = **0.019** (or 1.9%)
- Rule of thumb (Cohen, 1992):
  - **0.01 = small effect**
  - 0.06 = moderate effect
  - 0.14 = large effect
- **Cohen's d is more commonly used for independent samples**

SPSS: Paired t-test

- Question: *"Does level of stress decrease after exercise?"*
- Go to: **Analyze → Compare Means → Paired samples T test**
- Select (one-click) the two variables you want to compare. In this case it's "before exercise stress" and "after exercise stress"
- When both variables are highlighted move them into the box labelled "Paired Variables"
- Click "OK"

- PS: don't forget descriptive statistics and normality check…

Output

- We can see **stress score reduced after exercise**
- The "Paired samples test" table shows us that
  - The mean difference is **3.368**
  - **t = 52.760** which is significant **p=<.0001** (not .000)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	beforeexercisestress	26.73	433	5.848	.281
	afterexercisestress	23.36	433	6.002	.288

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	beforeexercisestress & afterexercisestress	433	.975	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	beforeexercisestress - afterexercisestress	3.368	1.329	.064	3.243	3.494	52.750	432	.000

Interpretation

- So difference is **highly significant** (reliable)
- What is the effect size?
- Eta squared = $t^2 / (t^2 + N – 1)$
- $52.75^2 / (52.75^2 + 433 – 1) = 0.866$

- Rule of thumb:
  - 0.01 = small effect
  - 0.06 = moderate effect
  - 0.14 = large effect
- **So this is a large effect!**

Non-parametric tests

- Parametric tests are quite robust to minor violations, particularly for reasonably large samples
- However, if major violations (e.g. **Highly skewed, ordinal data, inequality of variance**), to be safe:
  - run a non-parametric equivalent and compare the results
  - Pick the most conservative result (highest p)
- T-test equivalents (see Pallant, Chapter 16)
  - Independent samples - Mann-Whitney U
  - Paired samples - Wilcoxon Signed Ranks
- Correlation
  - Spearman’s rho (rank-order correlation)
  - Same procedure as Pearson, just check “Spearman” on the SPSS dialog
- Chi-square (see Pallant, Ch. 16)
  - For comparing frequencies between groups in one or more categories (non continuous variable)
  - E.g. Are men more likely to be smokers than women?
  - We’ll look at Chi-square in the next lab

Choice of parametric/non-parametric 2-sample test

	Independent Samples?	**Related Samples?**
Normal distribution and continuous data?	Independent samples t-test	Paired samples t-test
Non-normal distribution and/or ordinal data?	Mann-Whitney U-test	Wilcoxon test

On SPSS (v.20): Analyze → Nonparametric tests

On SPSS (older):

- Analyze → Nonparametric tests → 2 Independent samples → Tick Mann-Whitney U-test
- Analyze → Nonparametric tests → 2 Related samples → Tick Wilcoxon

Summary

- Descriptive statistics and graphs can tell us if a difference/relationship is evident
- Statistical tests like Pearson’s and T-test tell us if the relationship/difference is reliable (significance) and meaningful (effect size)
- Need to choose test carefully, particularly with respect to parametric assumptions
- In particular, non-parametric tests should be used for ordinal/nominal data or when distribution is far from normal

The data analysis process

- Prepare codebook
- Set up structure of data file
- Enter data
- Screen data file for errors
- Explore data using descriptive statistics and graphs
- Modify variables for further analyses
- Then either/both:
  - Conduct statistical analyses to explore relationships
  - Conduct statistical analyses to compare groups

Review this lecture

- Pallant, J. (2007/10)
  - Chapter 10 – Choosing the right statistic
  - Chapter 11 – Correlation
  - Chapter 17 – t-tests
  - Chapter 16 – non-parametric statistics
- If you don’t have a hard copy, Pallant (2010 version) can be accessed as an e-book accessible from the library catalogue