EDUS362 QUANTITATIVE RESEARCH METHOD

Demo 2.

Demo 2 Contents

1. State the hypotheses
2. Types of statistical technique
3. Transform variables
4. Crosstabs and Chi-square (χ^2) test
5. Create crosstabs in Excel

State the hypotheses

- Statistical hypotheses are assumption about population based on a collected sample. This assumption may or may not be true.
- Before you start analysis, you need to state a null hypothesis and an alternative hypothesis.

The hypotheses using above example are:

H_0: There is <u>no difference</u> between males and females on whether they smoke

H_1: There is <u>a difference</u> between males and females on whether they smoke

Rule for decision

- We use statistical analysis to test whether we accept or reject the hypothesis.

- Use *p*-value for the decision.

  p-value: the probability of observing our statistics if the null hypothesis is true.

- Statistically significant value for *p*-value needs to be .05 or smaller.

- Usually it is presented as $p < .05$(*), $p < .01$(**) or $p < .001$(***).

Type 1 and 2 Error

- There is always the possibility of reaching the wrong conclusion.

	H_0 is actually:	
	True	False
Reject H_0	Type 1 error	Correct
Accept H_0	Correct	Type 2 error

Two different types of statistical technique

1. Parametric statistics
   - This includes assumptions about the shape of the population (mainly normal distribution).
   - E.g. T-test, ANOVA (analysis of variance)

2. Non-parametric statistics
   - This do not make assumptions about the underlying population distribution (sometimes it is called distribution-free tests).
   - E.g. Chi-square (χ^2) test
   - Since it is less powerful than parametric tests, there may be possibilities to fail detecting differences between groups that actually exist.

→ It is better to use parametric statistics if you can.

Chi-square (χ^2) test for independence

- This test is used when you have **two categorical variables** from a single population. Each of these variables can have two or more categories. You can analyse whether there is significant relationship between them.

E.g. “Gender” and “Whether they smoke or not”

	Male	**Female**
Smoke	nnnn	nnnn
No smoke	nnnn	nnnn

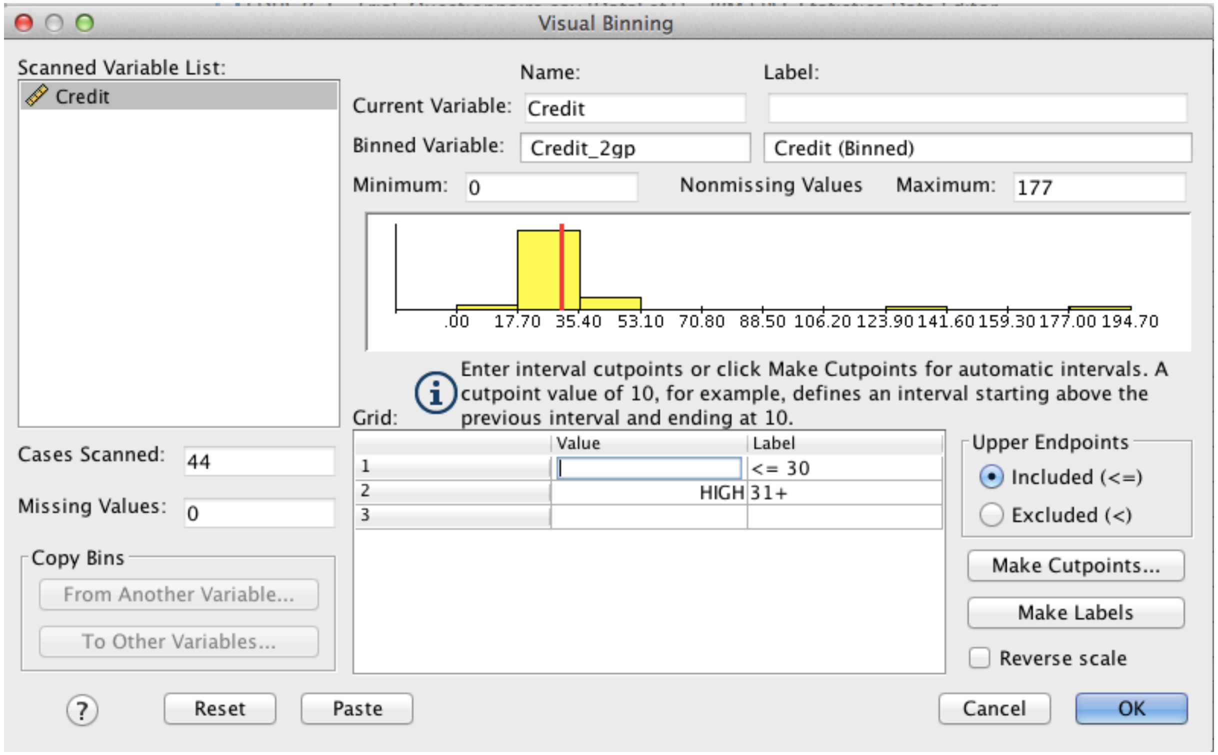
N.B.! The lowest expected frequency in each cell should be **5 or more**.

Exercise 1a: Transform variables

- Transform continuous variable (a number of credits) into categorical variable (two groups).

→ To conduct χ^2-test, we transform continuous variable (a number of credits) into categorical variable (two groups).

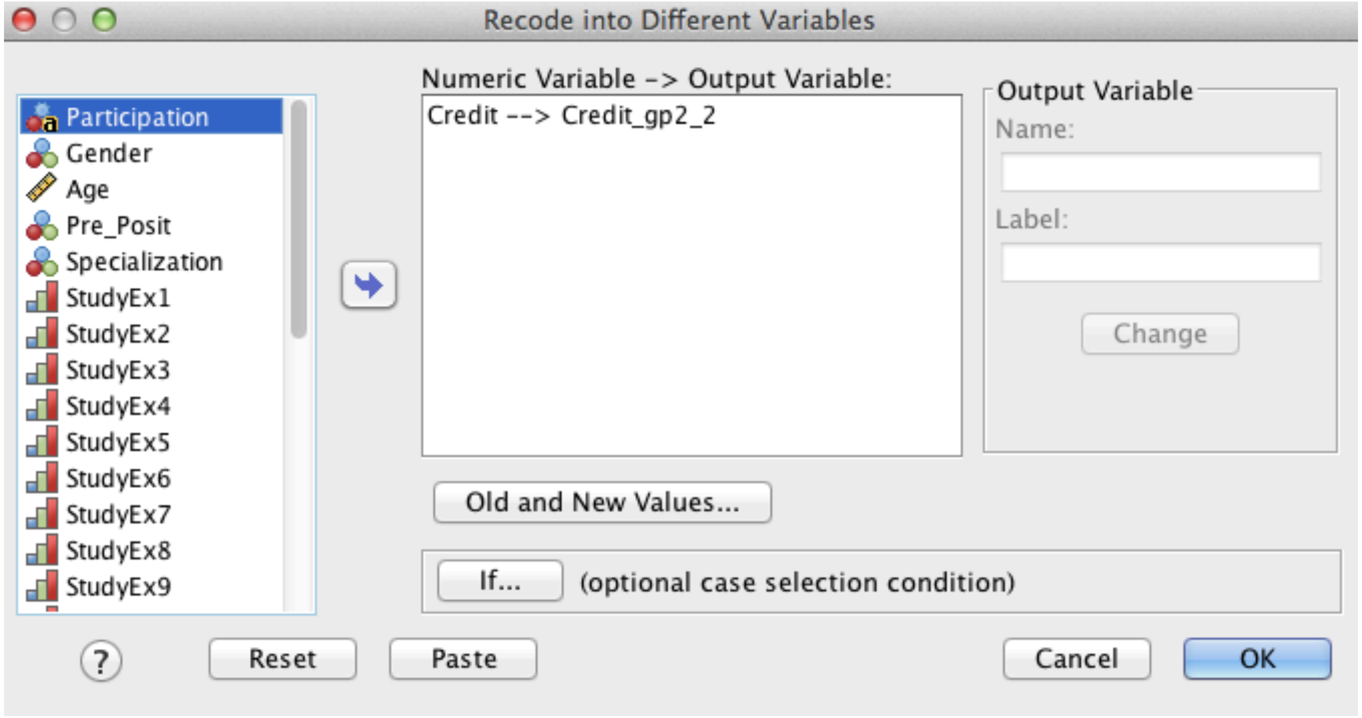
- Using “Visual Binning”.



Exercise 1b: Transform variables

- Transform continuous variable (a number of credits) into categorical variable (two groups).

- Recoding a categorical variable with your own criteria.



Independent and dependent variables

1. Dependent (or outcome) variables
   – This will vary according to the independent variable. The dependent variable is often the phenomenon which we are interested in and which is being studied.
   – E.g. political attitude, motivation, life satisfaction etc.

2. Independent variables
   – This will be investigated as possibly having a causal effect upon the dependent variables. It is often measured as a characteristic of the participant.
   – E.g. gender, occupation etc.

Exercise 2: Crosstabs and Chi-square (χ^2) test

- Variables are “Gender” and “Credit_gp2_1 ”
- Which is Independent and dependent variable?
- What is H_0 and H_1 in this case?

Exercise 2: Crosstabs and Chi-square (χ^2) test

Dependent variable

Independent variable

Credit (Binned) * Gender Crosstabulation

			Gender		
			Male	Female	Total
Credit (Binned)	Credit<31	Count	4	19	23
		% within Credit (Binned)	17.4%	82.6%	100.0%
		% within Gender	36.4%	57.6%	52.3%
		% of Total	9.1%	43.2%	52.3%
	31<=Credit	Count	7	14	21
		% within Credit (Binned)	33.3%	66.7%	100.0%
		% within Gender	63.6%	42.4%	47.7%
		% of Total	15.9%	31.8%	47.7%
Total		Count	11	33	44
		% within Credit (Binned)	25.0%	75.0%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	25.0%	75.0%	100.0%

Exercise 2: Crosstabs and Chi-square (χ^2) test

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.488[a]	1	.223		
Continuity Correction[b]	.759	1	.384		
Likelihood Ratio	1.498	1	.221		
Fisher's Exact Test				.303	.192
Linear-by-Linear Association	1.454	1	.228		
N of Valid Cases	44				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.25.

b. Computed only for a 2x2 table

→ other tables

→ 2 by 2 table

$X^2(1) = 0.76, p = .38$

Degrees of freedom

Value of test statistic

p-value

The null hypothesis is accepted because the *p*-value is not significantly small. ($p > .05$).

Exercise 2: Crosstabs and Chi-square (χ^2) test

- **Degree of freedom (*df*):** the number of values that are free to vary in the calculation.

  E.g. Three persons' height average is 175 cm. (It means the total is 525 cm.) **Two** persons' height can be whatever within human's height limit (e.g. "160 and 170", or "165 and 175"), <u>but third man's height is predetermined by two persons' height values.</u>

  df = 3-1 = **2**

e.g. 1	**160**	**170**	195	**175 (total 525)**
e.g. 2	**165**	185	**175**	**175 (total 525)**

*Degrees of Freedom (*df*) of crosstab

df = (number of cells in the row - 1) × (number of cells in the column - 1).

e.g. 2 × 2 table's *df* = (2 - 1) × (2 - 1) = **1**

Exercise 2: Crosstabs and Chi-square (χ^2) test

- Degrees of Freedom (*df*)

Table 10.7. **The chi-square values for different probability levels (.001–0.95), for degrees of freedom ranging from 1 to 50.**

df	Probabilities									
	.95	**.90**	**.70**	**.50**	**.30**	**.20**	**.10**	**.05**	**.01**	**.001**
1	.004	.016	.15	.46	1.07	1.64	2.71	3.84	6.64	10.83
2	.10	.21	.71	1.39	2.41	3.22	4.61	5.99	9.21	13.82
3	.35	.58	1.42	2.37	3.67	4.64	6.25	7.82	11.35	16.27
4	.71	1.06	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.15	1.61	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.64	2.20	3.83	5.35	7.23	8.56	10.65	12.59	16.81	22.46
7	2.17	2.83	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.13
9	3.33	4.17	6.49	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.87	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
11	4.58	5.58	8.15	10.34	12.90	14.63	17.28	19.68	24.73	31.26
12	5.23	6.30	9.03	11.34	14.01	15.81	18.55	21.03	26.22	32.91
13	5.89	7.04	9.93	12.34	15.12	16.99	19.81	22.36	27.69	34.53
14	6.57	7.79	10.82	13.34	16.22	18.15	21.06	23.69	29.14	36.12
15	7.26	8.55	11.72	14.34	17.32	19.31	22.31	25.00	30.58	37.70
20	10.85	12.44	16.27	19.34	22.78	25.04	28.41	31.41	37.57	45.32
25	14.61	16.47	20.87	24.34	28.17	30.68	34.38	37.65	44.31	52.62
30	18.49	20.60	25.51	29.34	35.53	36.25	40.26	43.77	50.89	59.70
50	34.76	37.69	44.31	49.34	54.72	58.16	63.17	67.51	76.15	86.66

http://www.eplantscience.com/index/genetics/linkage_and_crossing_over_in_diploid_organisms_higher_eukaryotes/chisquare_test.php

Exercise 2: Crosstabs and Chi-square (χ^2) test

- Effect size
  - Higher value indicating stronger association between the two variables.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-.184	.223
	Cramer's V	.184	.223
N of Valid Cases		44	

→ 2 by 2 table
→ other tables

Cohen's (1988) criteria for Phi coefficient:
Φ >.10 or Φ < -.10 : small effect
Φ >.30 or Φ < -.30 : medium effect
Φ >.50 or Φ < -.50 : large effect

N.B.! Cramer's V has different criteria.

Exercise 3: Report the result

- *Example*

A Chi-square test for independence indicated no significant (or significant) association between gender and smoking status, $\chi 2$ (1, n = 436) = .34, p = .56.

Exercise 4:
Create crosstabs in Excel

Table 1. Amount of credits between male and female

	Male	Female	Total
≦29 credit	3	6	9
	(8.6%)	(17.1%)	(25.7%)
≧30 credit	6	20	26
	(17.1%)	(57.1%)	(74.3%)
Total	9	26	35
	(25.7%)	(74.3%)	(100.0%)