Performance Analysis in Team games (ΑΠ-197επ)

Lecture 9

Managing categorical sports data

Categorical Data - Frequencies

- Sports Performance Analysis
  - Nominal variables
    - Nominal variables with neighbourhounding values (mapping of the court/pitch)
  - Ordinal variables
  - Frequency Data
    - Example (Attack directions of a volleyball player)
  - Comparison of observed - theoretical frequencies
    - chi-square goodness of fit
  - Comparison of observed – expected frequencies
    - chi-square test of independence

Chi-square goodness of fit (χ^2 goodness of fit)

- Chi-square Goodness-of-Fit Test checks the agreement between the observed frequencies of a qualitative variable and its theoretical frequencies
- Assumptions
  - H_0: There are no differences between the frequencies.

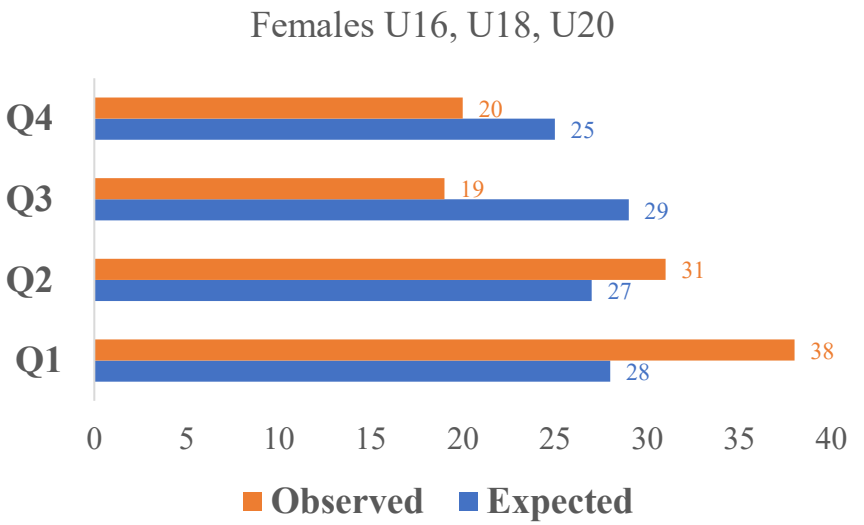
- **If *p* <.05**: rejection of H_0 There are differences between the frequencies.
- **If *p* >.05,** do not reject H_0. There are no differences between the frequencies.
  - Statistical significance indicates the probability that a result is not due to random sampling error
    - If a result is statistically significant, then it is likely to reflect a real effect on the population and not simply a random difference that occurred in the sample.
      - For a result to be considered statistically significant, its p-value must be less than a predetermined level (e.g. .05)
        - Statistical significance does not always mean that the result is practically or theoretically significant. A result may be statistically significant, but the effect may be so small that it has no practical impact on reality.

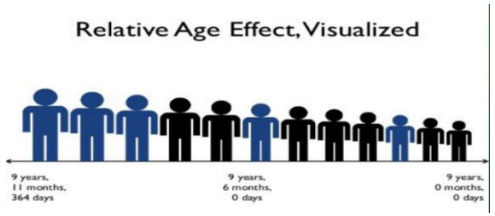
Example χ^2 goodness of fit

Data from the HBF

- Relative Age Effect (RAE)
  - Children born in the first months of the year excel physically and mentally and are overrepresented in selections
- Found in many sports.
- Basketball
- 4 Quarters/year (Q1-Q4)
- Females born 2003-2010
  - National teams U16, U18, U20
  - Total registrations in the Hellenic Basketball Federation (HBF)
  - Total births in Greece (data from H.S.A.)

FEMALES Quarter of birth				
	Observed N	Expected N	Expected N	Residual
Q1	38	27	28,0	10,0
Q2	31	27	26,8	4,2
Q3	19	27	28,5	-9,5
Q4	20	27	24,7	-4,7
Total	108			





Data from the thesis of Vicky Kyriakaki (2024) https://pergamos.lib.uoa.gr/uoa/dl/frontend/el/browse/3415089

Calculation steps for χ^2 goodness of fit

- Creation of the table for the observed frequencies(observed, O)
- Set the null hypothesis (eg. H_0 : the 2 factors are independent)
  - H_0: There are no differences between the frequencies.
  - **If *p* <.05**: rejection of H_0. There are differences in frequencies, so there is a dependence between selection and quarter of birth.
  - **If *p* >.05,** do not reject H_0. There are no differences between the frequencies, so there is no dependence between selection and quarter of birth.
- I calculate and tabulate the expected/theoretical frequencies(expected, E)
- For each cell I calculate the quantity $= \frac{(0-E)^2}{E}$
- Calculate the quantity $\chi^2 = \Sigma \frac{(0-E)^2}{E}$
- I identify the degrees of freedom: df= (c-1) (r-1), r=number of columns, c=number of rows
- Excel function CHISQ.DIST.RT. Calculate *p-value*.
  - Reject or do not reject H_0
  - EXCEL

Chi-square test of independence (χ^2 test of independence)

- Chi-square test of independence checks the independence of two qualitative variables
- Assumptions
  - H_0: There are no differences between the frequencies.

- **If *p* <.05**: rejection of H_0 There are differences between the frequencies.
- **If *p* >.05,** do not reject H_0. There are no differences between the frequencies.

Example χ^2 test of independence

- Volleyball
- Attack performance (A)
  - Lost Attack (LA)
  - Continued Attack (CA)
  - Won Attack (WA)
- In-game Role (R)
  - Outside Hitter (OH)
  - Middle Hitter (MH)
  - Opposite hitter (OPP)
  - Setter (S)
- If (A) και (R) are completely independent variables, then we expect the distribution of (A) in the entire sample (N=2434) to apply for each R separately.

	Outside Hitter	Middle Hitter	Opposite hitter	Setter	Total
Lost Attack	158	86	135	5	384
Continued Attack	484	200	343	24	1051
Won Attack	403	266	308	22	999
Total	1045	552	786	51	2434

Data from the thesis of Th. Sarakatsani (2023) https://pergamos.lib.uoa.gr/uoa/dl/frontend/el/browse/3335227

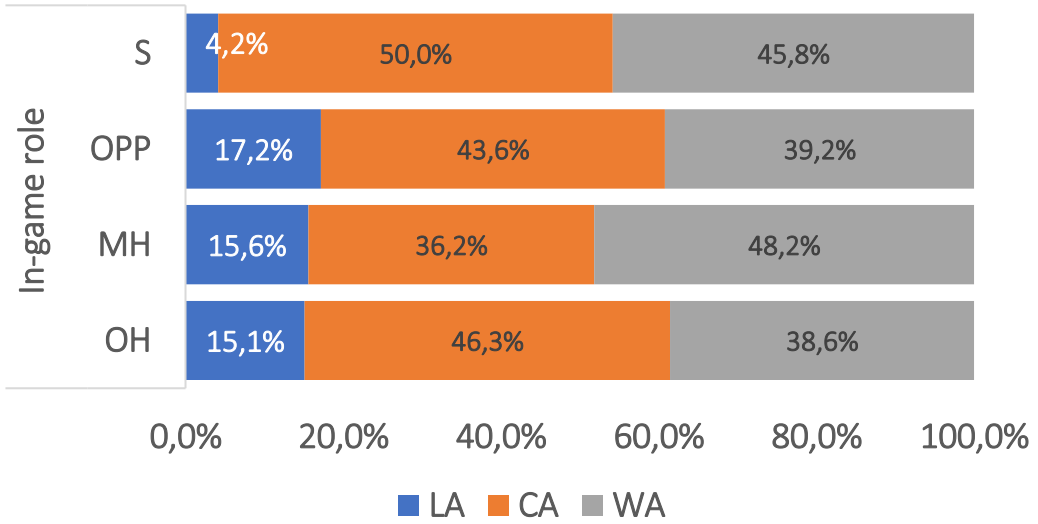
Calculation steps for χ^2 test of independence

- Creation of the table for the observed frequencies(observed, O)
- Set the null hypothesis (eg. H_0 : the 2 factors are independent)
  - H_0: There are no differences between the frequencies.
  - **If *p* <.05**: rejection of H_0. There are differences in frequencies, so there is a dependence between performance in Attack and in-game role of the player
  - **If *p* >.05,** do not reject H_0. There are no differences between the frequencies, so there is no dependence between performance in Attack and in-game role of the player
- I calculate and tabulate the expected frequencies(expected, E)
- For each cell I calculate the quantity $= \frac{(0-E)^2}{E}$
- Calculate the quantity $\chi^2 = \Sigma \frac{(0-E)^2}{E}$
- I identify the degrees of freedom: df= (c-1) (r-1), r=number of columns, c=number of rows
- Excel function CHISQ.DIST.RT. Calculate *p-value*.
  - Reject or do not reject H_0
  - EXCEL
- Additional items
  - Residuals (next slide, a statistical programme is neede to produce them)
  - (Effect size) Cramer's V (other course)

Calculation steps for χ^2 test of independence

Attack Performnce		Ειδικεύσεις				Σύνολο
		A	K	Δ	Π	
LA	Count	158	86	135	5	384
	%	15,1%	15,6%	17,2%	4,2%	15,7%
	Residuals	-,7	-,1	1,4	**-2,2**	
CA	Count	484	200	343	24	1051
	%	46,3%	36,2%	43,6%	50,0%	43,2%
	Residuals	**2,7**	**-3,8**	,3	1,0	
Residuals	Count	403	266	308	22	999
	%	38,6%	48,2%	39,2%	45,8%	41,1%
	Residuals	**-2,2**	**3,9**	-1,3	,7	
	Count	1045	552	786	51	2434
	% total	43,0%	22,7%	32,3%	2,0%	100,0%

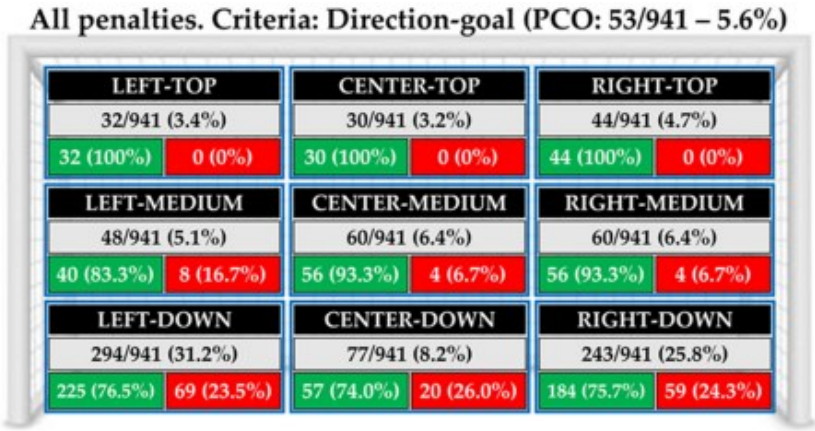
Residuals >2: More than expected
Residuals <-2: Less (fewer) than expected



Data from the thesis of Th. Sarakatsani (2023) https://pergamos.lib.uoa.gr/uoa/dl/frontend/el/browse/3335227

Soccer-Penalties Shoot-out

	Total	Goals	Saves
Left top	32	32	0
Center top	30	30	0
Right top	44	44	0
Left-medium	48	40	8
Center-medium	60	56	4
Right-medium	60	56	4
Left down	294	225	69
Center-down	77	57	20
Right-down	243	184	59
Post or outside	53		
Total	941	724	164



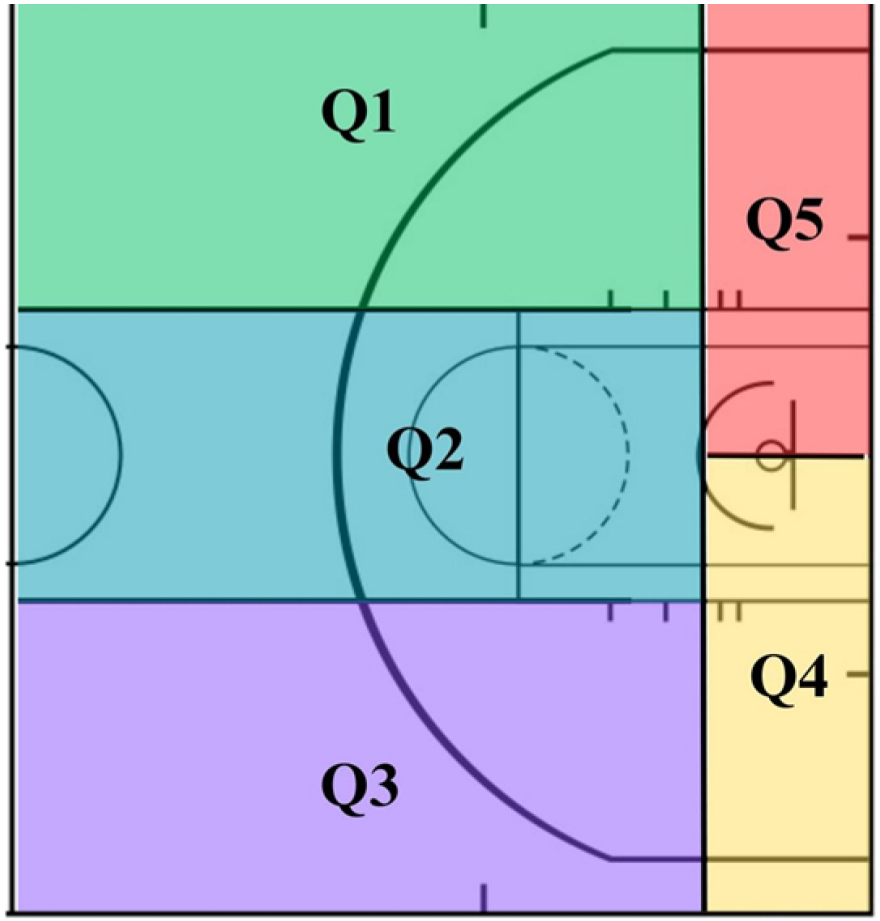
1. Is there a dependence between the shooting area and the selection of the players?
2. Is there a dependence between the shooting area and the effectiveness?

https://www.mdpi.com/2076-3417/14/16/7046

Basketball-Ending court zone

	Q1	Q2	Q3	Q4	Q5
Basket	45	49	39	37	25
No Basket	16	50	14	42	36

1. Is there a dependence between the shooting area and the selection of the players?
2. Is there a dependence between the shooting area and the effectiveness?



https://www.mdpi.com/2076-3417/14/23/10883

Basketball-Ending court zone

	A2	A3	B2	B3
LA	45	46	24	24
CA	99	111	56	49
WA	250	221	91	73

1. Is there a dependence between the setting area and the attack performance?

Data from the thesis of Ch. Salonidou (2023) https://pergamos.lib.uoa.gr/uoa/dl/frontend/el/browse/3336294

Performance Analysis in Team games (ΑΠ-197επ)

Lecture 9

Managing categorical sports data