

## Z-statisztika

- Amely (a 0-hipotézis teljesülése esetén) jó közelítéssel standardizált normális eloszlást ( $\mu=0, \sigma=1$ ) követ
- Általában valamilyen más mintaváltozó standardizálásával számoljuk:

$$Z = \frac{(\text{eredeti változó}) - (\text{várható értéke})}{\text{[a hibája (SE)}}$$

- Alkalmazások:

– részarányok és különbségük

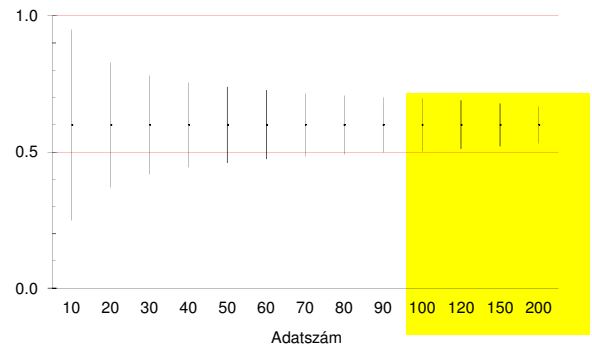
– beütésszámok különbsége :  $Z = \frac{N_2 - N_1}{\sqrt{N_1 + N_2}}$

Varga J.

Statisztika: gyakori módszerek és hibák

60

## Részarány konfidencia-intervalluma



Varga J.

Statisztika: gyakori módszerek és hibák

61

## Részarányok összehasonlítása: Excel

Pl.: Infarktusos betegek halálzási aránya nitrát adásakor és anélkül (Chiche et al.)

Tapasztalt gyakoriságok:	Meghalt	Túlélő	Összesen
Nitrátadás mellett	3	47	50
Kontroll	8	37	45
Összesen:	11	84	95

Varga J.

Statisztika: gyakori módszerek és hibák

62

## Részarányok összehasonlítása: SPSS

Pl.: Infarktusos betegek halálzási aránya nitrát adásakor és anélkül (Chiche et al.)

Kezelés \* Kimenetel Crosstabulation

Kezelés	Kimenetel	Kimenetel		Total
		Túlélő	Meghalt	
Kontroll	Nitrát	37	8	45
Total		47	3	50
		84	11	95

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.209 <sup>a</sup>	1	.073		
Continuity Correction <sup>b</sup>	2.162	1	.141		
Likelihood Ratio	3.288	1	.070		
Fisher's Exact Test				.108	.070
Linear-by-Linear Association	3.175	1	.075		
N of Valid Cases	95				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.21.

Varga J.

Statisztika: gyakori módszerek és hibák

63

## Esélyarány (odds ratio)

Mantel-Haenszel Common Odds Ratio Estimate

Estimate		.295
ln(Estimate)		-1.220
Std. Error of ln(Estimate)		.712
Asymp. Sig. (2-sided)		.087
Asymp. 95% Confidence Interval	Common Odds Ratio	.073
	Lower Bound	1.191
	Upper Bound	
	ln(Common Odds Ratio)	-2.615
	Lower Bound	
	Upper Bound	.175

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

Varga J.

Statisztika: gyakori módszerek és hibák

64

## Részarányok összehasonlítása

Pl.: Infarktusos betegek halálzási aránya nitrát adásakor és anélkül (Whitley and Ball, Critical Care 6/3: 222-225)

Trial	Intravenous nitrate	Control
Chiche	3/50	8/45
Bussman	4/31	12/29
Flaherty	11/56	11/48
Jaffe	4/57	2/57
Lis	5/64	10/76
Jugdutt	24/154	44/156

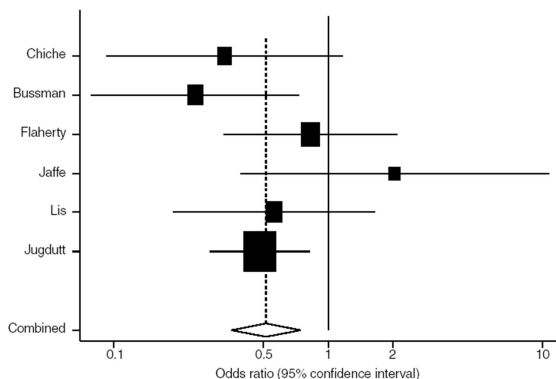
Numbers: dead/randomized

Varga J.

Statisztika: gyakori módszerek és hibák

65

## Esélyarányok konfidencia-intervallumokkal



Varga J.

Statisztika: gyakori módszerek és hibák

66

## Részarányok: mikor melyik próbát?

- 2\*2-es táblázatra: **Fisher-féle egzakt próba**
  - Kivéve a nagyon nagy esetszámot
  - Nem igényel folytonossági korrekciót
- $\chi^2$ -próba feltétele, hogy a várt érték:
  - a cellák < 1/5-ében van 5 alatt
  - Egyetlen cellában sincs 1 alatt
  - Nem nagyon nagy esetszámmal folytonossági korrekció (Yates)
- Ha nem, cellák **összeolvasztása**

Varga J.

Statisztika: gyakori módszerek és hibák

67

## Eset táblázat („contingency table”)

- Független egyedek/esetek?
- Nem párosíthatók az adatok?
- Tényleg esetszámok? (nem: arányok, gyak. eloszlások, ...)
- Trendanalízishez: a sorok/oszlopok természetes sorrendben vannak?

Varga J.

Statistika: gyakori módszerek és hibák

68

## Eset táblázatok kialakítása

- **Keresztmetszeti tanulmány:**
  - Először összeállítunk **egyetlen** betegcsoportot, aztán besoroljuk őket 2 szempont szerint
  - Pl.: ki volt-e téve nagyfeszültségnek; van-e agydaganata
  - NEM úgy, hogy kiválasztunk nagyfeszültségnek kitett egyéneket, **aztán** megnézzük, van-e agydaganata.
- **Prospektív tanulmány:**
  - Kiválasztunk 2 csoportot a **hatás szerint** (nagyfesz. nek kitett és nem kitett),
  - majd figyeljük bennük az agydaganat kialakulását

Varga J.

Statistika: gyakori módszerek és hibák

69

## Eset táblázatok kialakítása – folyt.

- **Visszatekintő (retrospektív) tanulmány:**
  - Összeállítunk 2 csoportot a **KIMENETEL** szerint (van-e agydaganata)
  - Aztán megnézzük bennük a hatás (nagyfeszültség) előfordulását
- **Kísérlet:**
  - Összeállítjuk egyedek egyetlen csoportját
  - A fele kezelést kap, a másik fele másféle vagy semmilyen kezelést
- **Diagnosztikai eljárás hatékonyságának lemerése:**
  - Két csoportot választunk: az egyikben jelen van a betegség/állapot, a másikban nem
  - Minden egyeden elvégezzük a vizsgálatot.

Varga J.

Statistika: gyakori módszerek és hibák

70

## Érzékenység és fajlagosság

Jelölés:

		Teszt (T):	
		+	-
Betegség (D):	Van	D+, T+	D+, T-
	Nincs	D-, T+	D-, T-

- Feltételes valószínűség:  $P(A|B)$  („A’ ha ‘B’”) mi a vals.-e, hogy B bekövetkezése esetén A is bekövetkezik
- Érzékenység (sensitivity) =  $P(T+|D+)$
- Fajlagosság (specificity) =  $P(T-|D-)$
- Pozitív prediktív érték =  $P(D+|T+)$
- Negatív prediktív érték =  $P(D-|T-)$

Varga J.

Statistika: gyakori módszerek és hibák

71

## Bayes-tétel:

A feltételes valószínűség megfordításának tétele

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{\sum_{j=1}^n P(B | A_j) \cdot P(A_j)}$$

Pl.: a pozitív prediktív értékre:

$$P(D+ | T+) = \frac{P(T+ | D+) \cdot P(D+)}{P(T+ | D-) \cdot P(D-) + P(T+ | D+) \cdot P(D+)}$$

Példa: Érzékenység =  $P(T+ | D+) = 80\%$   
Fajlagosság =  $P(T- | D-) = 95\%$

$$P(D+) = 25\% \quad P(D-) = 75\% \quad P(D+) = 5\% \quad P(D-) = 95\%$$

$$P(D+ | T+) = \frac{0.8 \cdot 0.25}{0.05 \cdot 0.75 + 0.80 \cdot 0.25} = 0.84 \quad P(D- | T-) = \frac{0.8 \cdot 0.05}{0.05 \cdot 0.95 + 0.80 \cdot 0.05} = 0.46$$

Varga J.

Statistika: gyakori módszerek és hibák

72

## Tanulság:

- Az érzékenység és a fajlagosság magas értéke nem jelent automatikusan nagy pontosságot
- Különösen a szűrővizsgálat esetét kell átgondolni (alacsony betegség-előfordulás).
- Ráadásul a saját módszer ill. értékelés érzékenysége és fajlagossága rosszabb lehet, mint az irodalomban leközölt.
- Az előfordulási gyakoriságnak arra a populációra kell vonatkoznia, amelyet vizsgálunk!  
Pl.: előválogatott betegcsoport

Varga J.

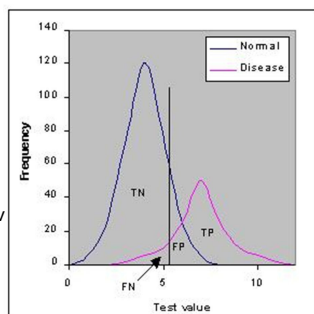
Statistika: gyakori módszerek és hibák

73

## Diagnosztikai határ szerepe

		Teszt:	
		+	-
Betegség:	Van	TP	FN
	Nincs	FP	TN

TP: „true pos.” valódi pozitív  
FN: „false neg.” álnegatív  
TN: „true neg.” valódi negatív  
FP: „false pos.” álpozitív



<http://gim.unmc.edu/dxtests/roc1.htm>

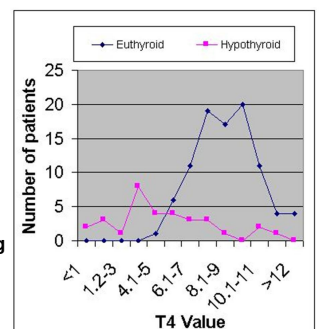
Varga J.

Statistika: gyakori módszerek és hibák

74

## Pl.: T<sub>4</sub> eu- és hypothyroid egyéneknél

T <sub>4</sub> érték	Hypothyroid	Euthyroid
≤ 5	18	1
5.1 - 7	7	17
7.1 - 9	4	36
≥ 9	3	39
<b>Összesen:</b>	<b>32</b>	<b>93</b>
<b>Határpont</b>	<b>Érzékenység</b>	<b>Fajlagosság</b>
5	0.56	0.99
7	0.78	0.81
9	0.91	0.42



Varga J.

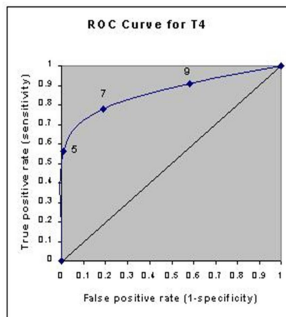
Statistika: gyakori módszerek és hibák

75

## ROC (Receiver Operating Characteristic curve)

T4 érték	Hypothyroid	Euthyroid
≤ 5	18	1
5.1 - 7	7	17
7.1 - 9	4	36
≥ 9	3	39
<b>Összesen:</b>	<b>32</b>	<b>93</b>

Határpont	Érzékenység	Fajlagosság
5	0.56	0.99
7	0.78	0.81
9	0.91	0.42



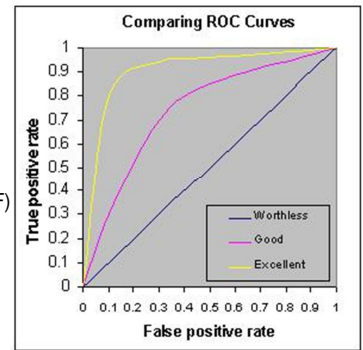
Varga J.

Statistika: gyakori módszerek és hibák

76

## ROC görbe alatti terület

- 0.90-1 = kiváló (A)
- 0.80-0.90 = jó (B)
- 0.70-0.80 = megfelelő (C)
- 0.60-0.70 = gyenge (D)
- 0.50-0.60 = nem használható (F)

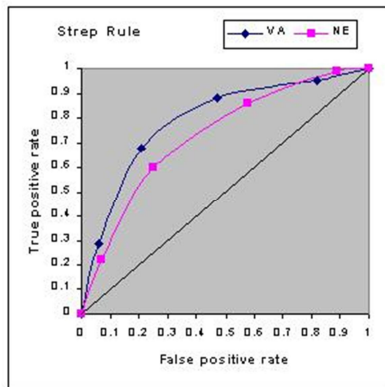


Varga J.

Statistika: gyakori módszerek és hibák

77

## Módszerek összehasonlítása



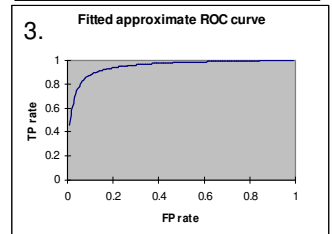
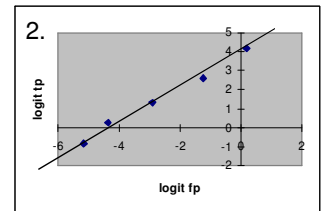
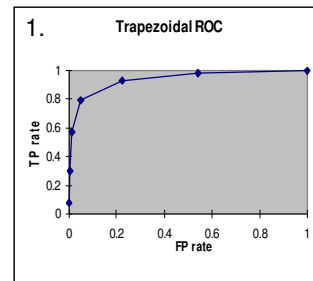
Varga J.

Statistika: gyakori módszerek és hibák

78

## ROC: Transzformáció és illesztés

$$\text{logit}(p) = \ln \left[ \frac{p}{1-p} \right]$$



Varga J.

Statistika: gyakori módszerek és hibák

79

## Tanulság

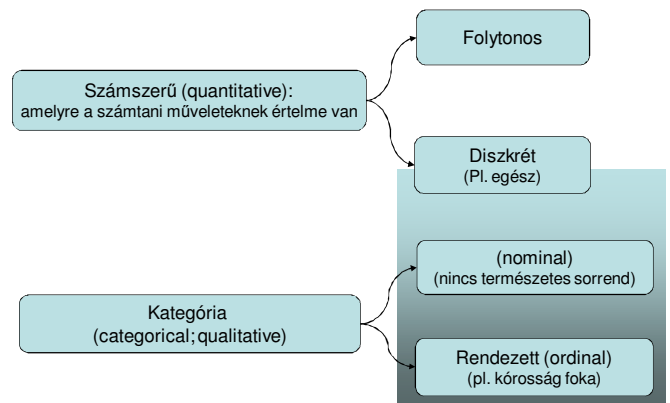
- Két számszerű diagnosztikai mutató hatékonyságának összehasonlítására a ROC-görbe használható
- A görbe alatti terület méri a „jóságot”
- „logit” transzformációval „kiegyenesíthető” a ROC görbe

Varga J.

Statistika: gyakori módszerek és hibák

80

## Adattípusok



Varga J.

Statistika: gyakori módszerek és hibák

81

## Eltérő kezelés

- Rendezett kvalitatív változó:  
NEM számolható belőle átlag, szórás,  
csak **részarányok**
- Az arányszámok, mint folytonos változók csak  
*arc sin* transzformáció után hasonlíthatók össze  
(mert a 0 és 1 közeli értékek kevésbé változékonyak)

Varga J.

Statistika: gyakori módszerek és hibák

82