

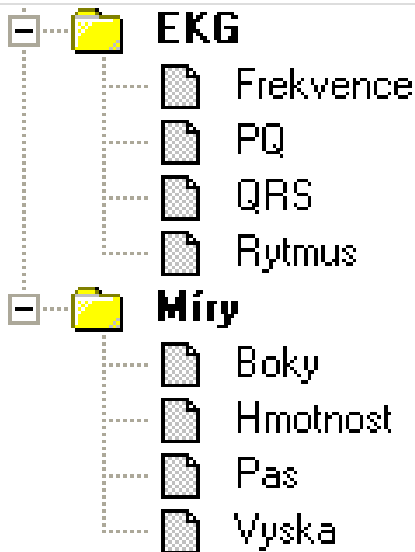
Vliv tělesných měr na EKG

Hedvika Peroutková

Matěj Cáha

Popis vytvořených atributů

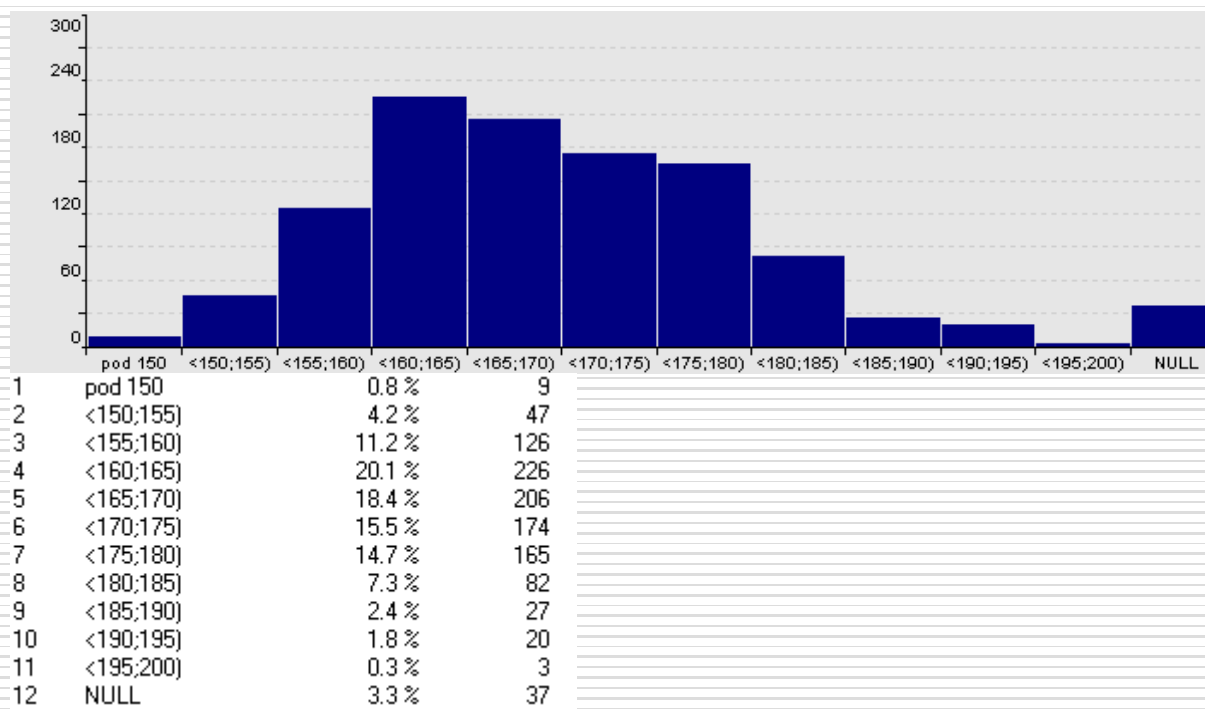
Zdatabaze „ADAMEK“, která poskytuje EuroMISE centrum (<http://www.euromise.cz/>), Oddělení medicínské informatiky Ústavu Informatiky AV ČR, Praha pro studijní účely, byly vytvořeny 2 skupiny atributů: EKG a Míry.



Míry - Výška

Atribut byl vytvořen ze sloupce „Výška“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 5cm.

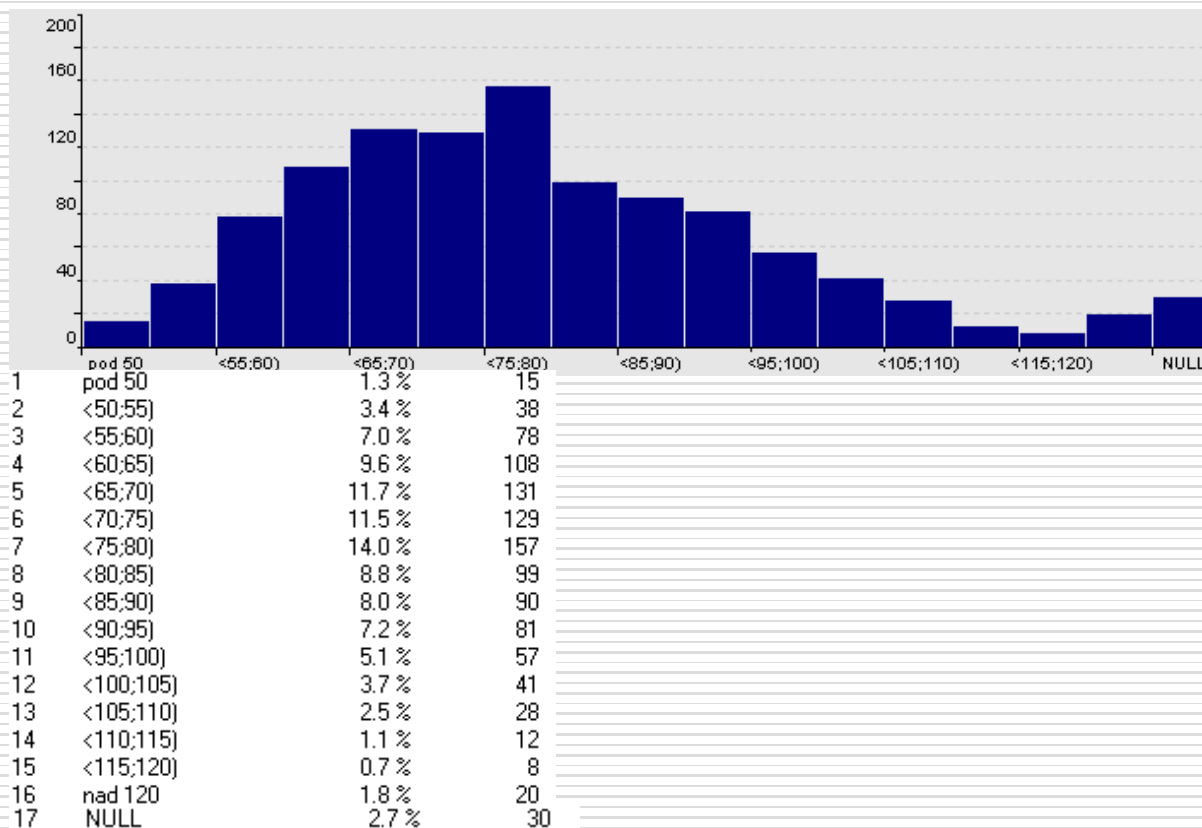
•počet kategorií : 11



Míry - Hmotnost

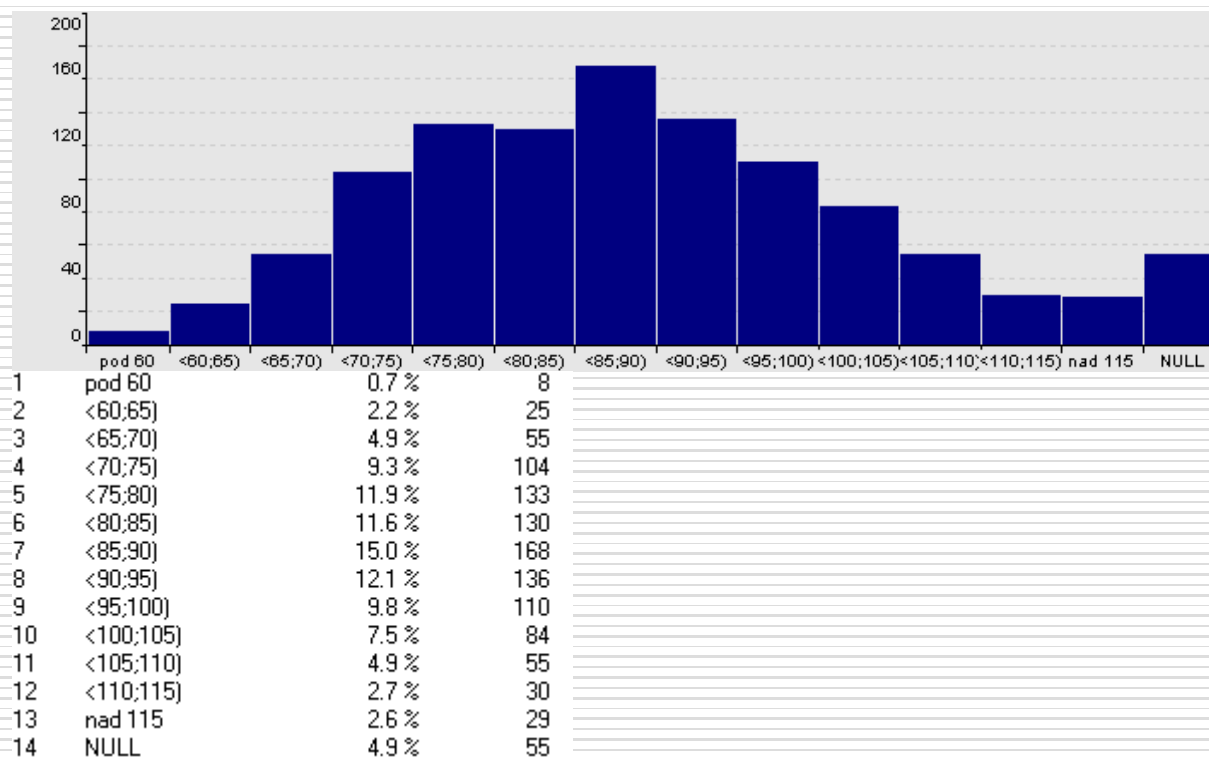
Atribut byl vytvořen ze sloupce „Hmotnost“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 5kg.

•počet kategorií : 16



Míry - Pas

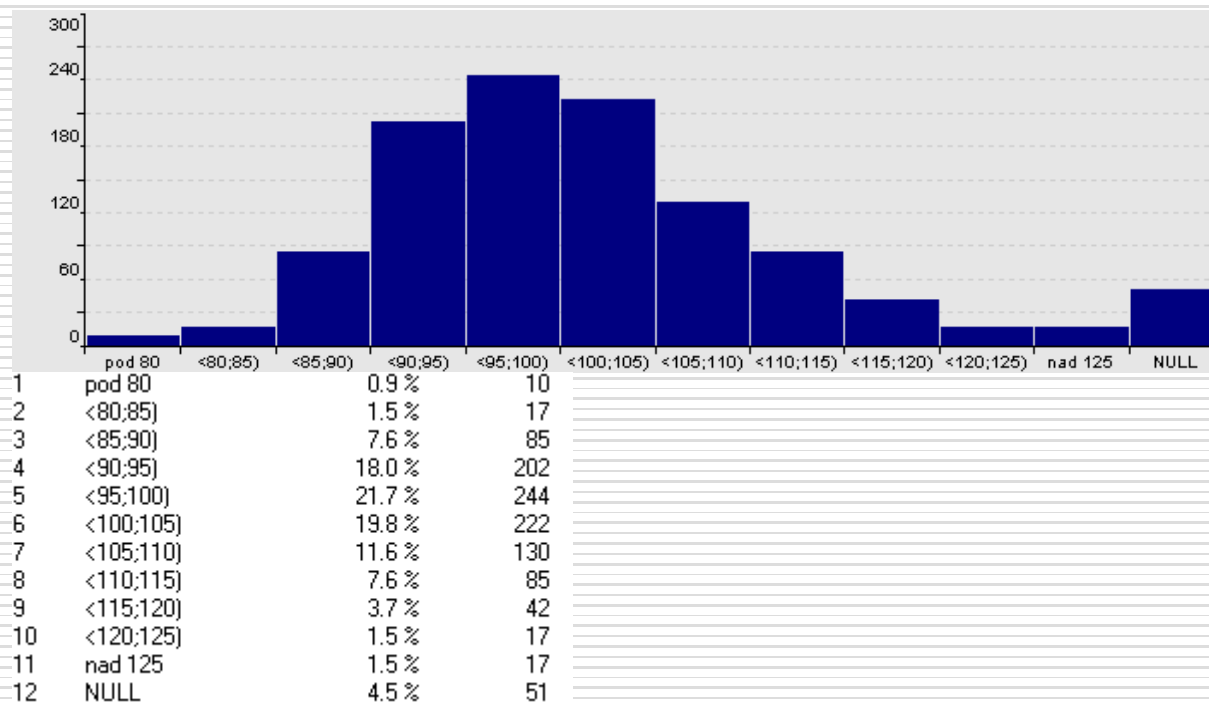
Atribut byl vytvořen ze sloupce „Pas“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 5cm .
•počet kategorií : 13



Míry - Boky

Atribut byl vytvořen ze sloupce „Boky“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 5cm.

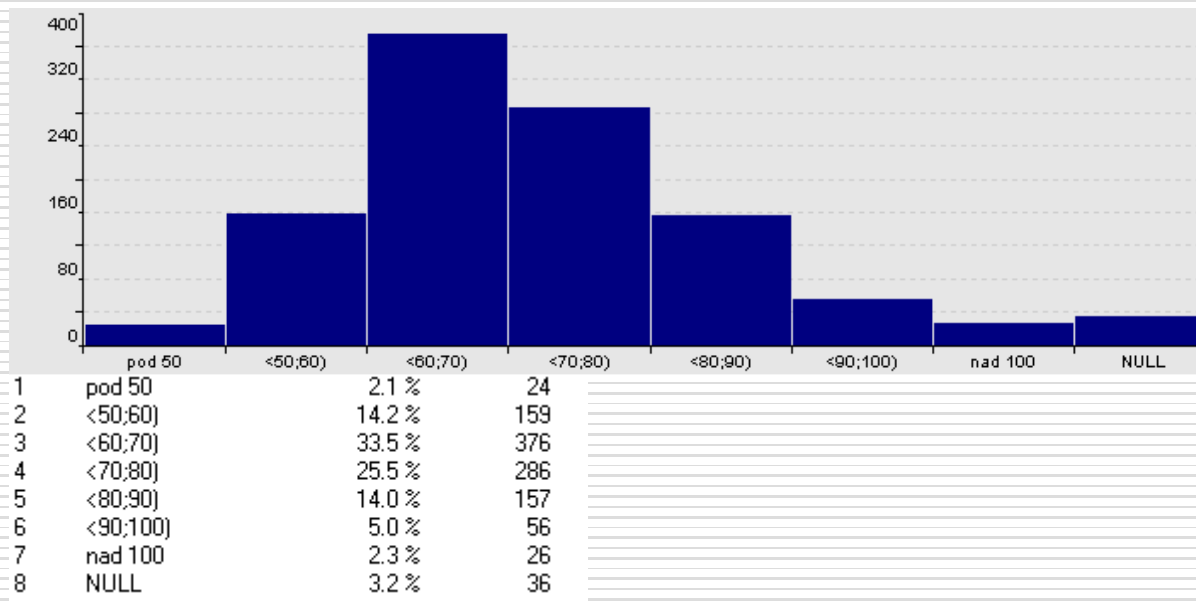
•počet kategorií : 11



EKG - Frekvence

Atribut byl vytvořen ze sloupce „Frek“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 10.

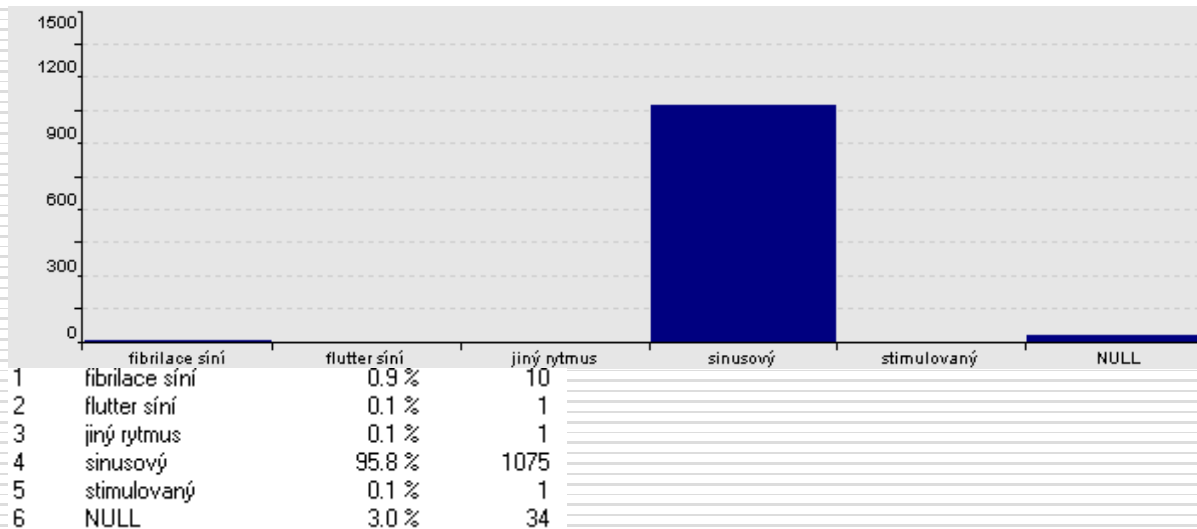
•počet kategorií : 7



EKG - Rytmus

Atribut byl vytvořen ze sloupce „Rytmus“ pomocí *Each value – one category*.

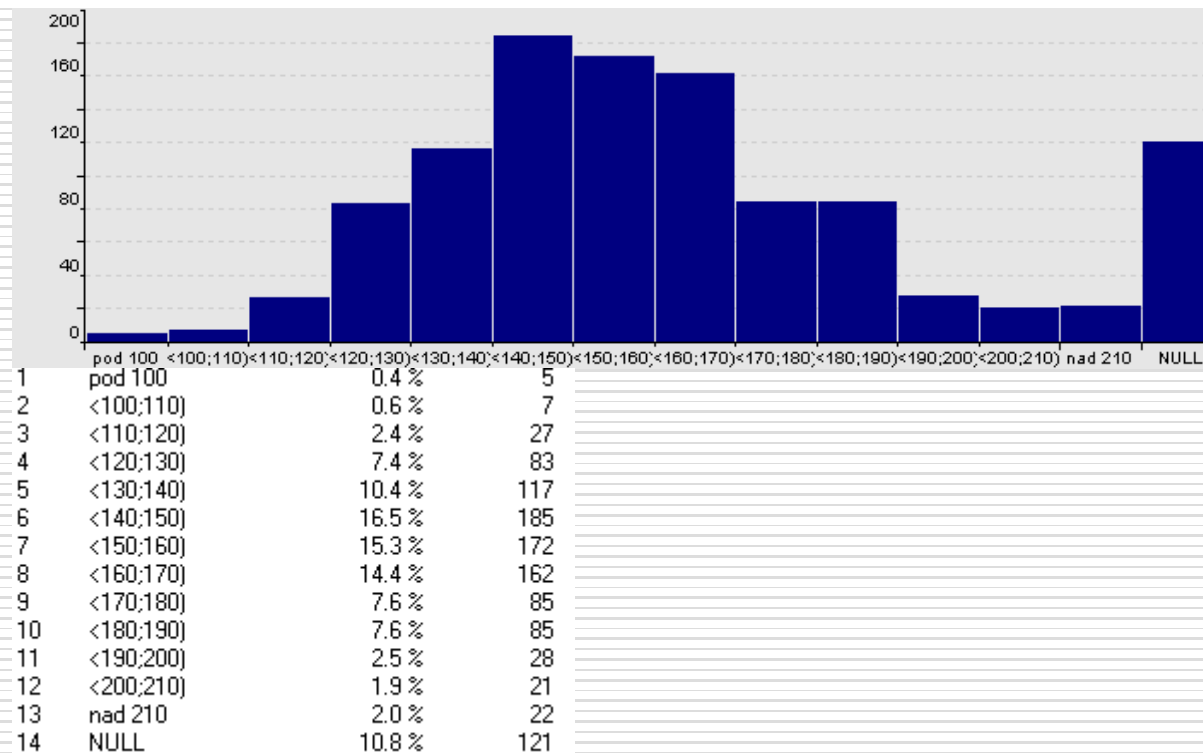
•počet kategorií : 5



EKG - PQ

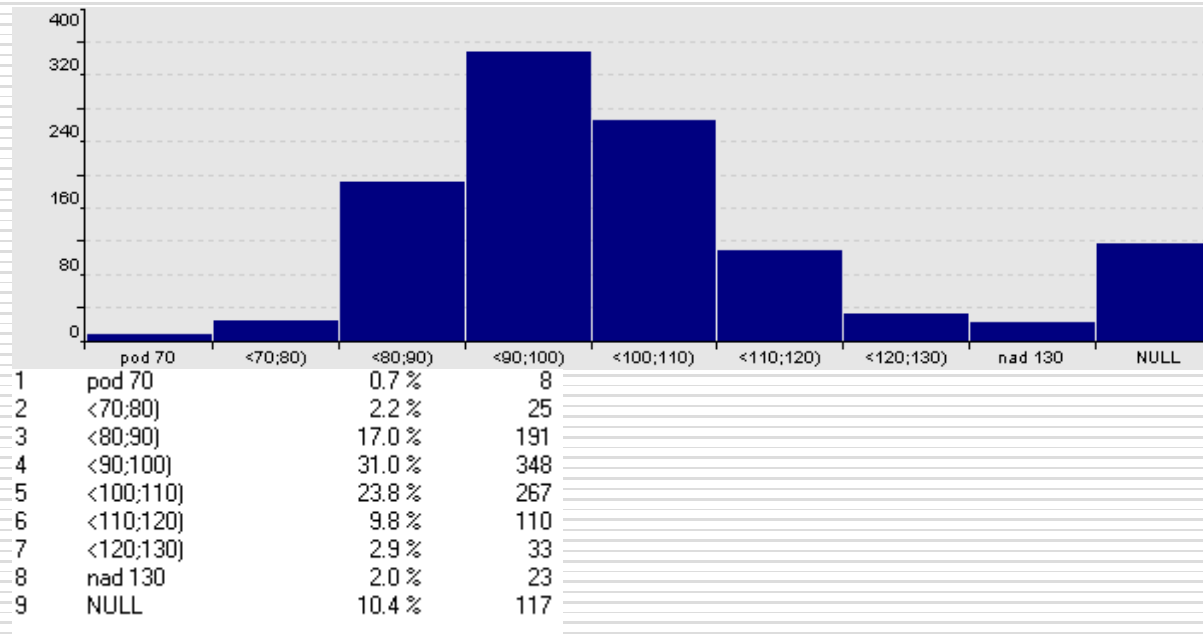
Atribut byl vytvořen ze sloupce „PQ_int_novy“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 10.

•počet kategorií : 13



EKG - QRS

Atribut byl vytvořen ze sloupce „QRS_int“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 10 .
•počet kategorií : 8



Zajímavé vztahy platící v datech

- o Zajímavé vztahy tvaru $\varphi \Rightarrow_{p; \text{Base}} \psi$ platící v datech (fundovaná implikace)
 - o Zajímavé vztahy tvaru $\varphi \Rightarrow_{+p; \text{Base}} \psi$ platící v datech (kladná odchylka od průměru)
-

Zajímavé vztahy tvaru $\varphi \Rightarrow_{p;Base} \psi$ platící v datech

(fundovaná implikace)

- Za φ jsme volili booleovský atribut odvozený ze skupiny atributů *Míry* a za ψ booleovský atribut odvozený ze skupiny atributů EKG.
- Zpočátku byly voleny za $P \geq 0.9$ a $BASE \geq 15$, tyto hodnoty byly upravovány pro dosažení optimálního počtu nalezených vztahů. Dalším parametrem bylo nastavování délky řezů u succedentů. Výsledné délky byly stanoveny tak, že kratší délky řezu již nedávaly žádné výsledky a větší již většinou nedávaly zajímavé výsledky.

	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

$$\Rightarrow_{0.9,15} (a,b,c,d) = \text{true}$$

$$\text{právě když } \frac{a}{a+b} \geq 0.9 \wedge a \geq 15$$

1) *Fundovaná implikace*

- *Antecedent 0-3 atributů* : Atribut Boky, Hmotnost, Pas, Výška – intervaly délky 1-3
 - *Succedent* : Atribut QRS - Řez délky 1 - 4
 - $P \geq 0.95$ a $BASE \geq 20$
 - *Výsledek* : Nalezeny 2 vztahy
-

1) Fundovaná implikace

Antecedent	Boky	<90;95], <95;100]
	Hmotnost	(<55;60]...<65;70])
	Pas	<65;70]
Succedent	QRS	(<= <90;100])

Contingency table

	Succedent	NOT Succedent	
Antecedent	22	1	23
NOT Antecedent	540	428	968
	562	429	991

Values from contingency table:

a	22	22	a-frequency from the contingency table
b	1	1	b-frequency from the contingency table
c	540	540	c-frequency from the contingency table
d	428	428	d-frequency from the contingency table
r	23	23	r-frequency (a+b) from the contingency table
n	991	991	n-frequency (a+b+c+d) from the contingency table
Conf	0.96	0.9565217391	Confidence (validity): $a/(a+b)$
DConf	0.04	0.0390763766	D-Confidence: $a/(a+b+c)$
EConf	0.45	0.454086781	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.02	0.0221997982	Support: $a/(a+b+c+d)$
Cmplt	0.04	0.0391459075	Completeness: $a/(a+c)$
AvgDf	0.69	0.6866780133	Average difference: $a(a+b+c+d)/((a+b)(a+c)) - 1$

1) Fundovaná implikace

Antecedent	Boky	<90;95), <95;100]
	Hmotnost	<55;60), <60;65)
	Paš	<65;70)
Succedent	QRS	(<= <90;100))

Contingency table

	Succedent	NOT Succedent	
Antecedent	21	1	22
NOT Antecedent	543	429	972
	564	430	994

Values from contingency table:

a	21	21	a-frequency from the contingency table
b	1	1	b-frequency from the contingency table
c	543	543	c-frequency from the contingency table
d	429	429	d-frequency from the contingency table
r	22	22	r-frequency (a+b) from the contingency table
n	994	994	n-frequency (a+b+c+d) from the contingency table
Conf	0.95	0.9545454545	Confidence (validity): $a/(a+b)$
DConf	0.04	0.0371681416	D-Confidence: $a/(a+b+c)$
EConf	0.45	0.4527162978	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.02	0.0211267606	Support: $a/(a+b+c+d)$
Cmplt	0.04	0.0372340426	Completeness: $a/(a+c)$
AvgDf	0.68	0.6823017408	Average difference: $a(a+b+c+d)/((a+b)(a+c)) - 1$

1) *Fundovaná implikace*

- *Antecedent 0-3 atributů* : Atribut Boky, Hmotnost, Pas, Vyska – intervaly délky 1-3
 - *Succedent* : Atribut PQ - Řez délky 1 – 7
 - $P \geq 0.95$ a $BASE \geq 25$
 - *Výsledek* : Nalezen 1 vztah
-

1) Fundovaná implikace

Antecedent	Pas	pod 60, <60;65)
	Vyska	(<155;160)...<165;170))
Succedent	PQ	(<= <150;160))

Contingency table

	Succedent	NOT Succedent	
Antecedent	26	0	26
NOT Antecedent	561	398	959
	587	398	985

Values from contingency table:

a	26	26	a-frequency from the contingency table
b	0	0	b-frequency from the contingency table
c	561	561	c-frequency from the contingency table
d	398	398	d-frequency from the contingency table
r	26	26	r-frequency (a+b) from the contingency table
n	985	985	n-frequency (a+b+c+d) from the contingency table
Conf	1	1	Confidence (validity): $a/(a+b)$
DConf	0.04	0.0442930153	D-Confidence: $a/(a+b+c)$
EConf	0.43	0.4304568528	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.03	0.0263959391	Support: $a/(a+b+c+d)$
Cmplt	0.04	0.0442930153	Completeness: $a/(a+c)$
AvgDf	0.68	0.6780238501	Average difference: $a(a+b+c+d)/((a+b)(a+c)) - 1$

1) Fundovaná implikace

- *Antecedent 0-2 atributů* : Atribut Boky, Hmotnost, Pas, Vyska – intervaly délky 1-3
 - *Succedent* : Atribut PQ - Řez délky 1 – 7
 - $P \geq 0.9$ a $BASE \geq 20$
 - *Výsledek* : Nalezeno 7 vztahů (jeden z nich je předcházející vztah), další dva nejzajímavější z hlediska hodnoty P jsou uvedeny níže.
-

1) Fundovaná implikace

Antecedent	Pas	<60;65)
	Vyska	(<155;160)...<165;170))
Succedent	PQ	(<= <150;160))

Contingency table

	Succedent	NOT Succedent	
Antecedent	20	0	20
NOT Antecedent	567	398	965
	587	398	985

Values from contingency table:

a	20	20	a-frequency from the contingency table
b	0	0	b-frequency from the contingency table
c	567	567	c-frequency from the contingency table
d	398	398	d-frequency from the contingency table
r	20	20	r-frequency (a+b) from the contingency table
n	985	985	n-frequency (a+b+c+d) from the contingency table
Conf	1	1	Confidence (validity): $a/(a+b)$
DConf	0.03	0.0340715503	D-Confidence: $a/(a+b+c)$
EConf	0.42	0.4243654822	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.02	0.0203045685	Support: $a/(a+b+c+d)$
Cmplt	0.03	0.0340715503	Completeness: $a/(a+c)$
AvgDf	0.68	0.6780238501	Average difference: $a(a+b+c+d)/((a+b)(a+c))- 1$

1) Fundovaná implikace

Antecedent	Boky	(<= <85;90))
	Pas	pod 60, <60;65)
Succedent	PQ	(<= <150;160))

Contingency table

	Succedent	NOT Succedent	
Antecedent	24	1	25
NOT Antecedent	562	387	949
	586	388	974

Values from contingency table:

a	24	24	a-frequency from the contingency table
b	1	1	b-frequency from the contingency table
c	562	562	c-frequency from the contingency table
d	387	387	d-frequency from the contingency table
r	25	25	r-frequency (a+b) from the contingency table
n	974	974	n-frequency (a+b+c+d) from the contingency table
Conf	0.96	0.96	Confidence (validity): $a/(a+b)$
DConf	0.04	0.0408858603	D-Confidence: $a/(a+b+c)$
EConf	0.42	0.4219712526	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.02	0.0246406571	Support: $a/(a+b+c+d)$
Cmplt	0.04	0.0409556314	Completeness: $a/(a+c)$
AvgDf	0.6	0.5956313993	Average difference: $a(a+b+c+d)/((a+b)(a+c)) - 1$

Zajímavé vztahy tvaru $\varphi \Rightarrow_{+p;Base} \psi$ platící v datech

(kladná odchylka od průměru)

- Za φ jsme volili booleovský atribut odvozený ze skupiny atributů *Míry* a za ψ booleovský atribut odvozený ze skupiny atributů EKG.
- *Antecedent 0-3 atributů* : Atribut Boky, Hmotnost, Pas, Vyska – intervaly délky 1-3
- *Succedent* : Atribut PQ - Řez délky 1 – 6
- $P \geq 1.5$ a $BASE \geq 15$
- *Výsledek* : Nalezeno 6 vztahů, nejzajímavější z hlediska odchylky od průměru (AvgDf) uveden níže

$$\varphi \Rightarrow_{+p,Base}^+ \psi \quad \frac{a}{a+b} \geq (1+p) \frac{a+c}{a+b+c+d} \wedge a \geq Base$$

2) Kladná odchylka od průměru

Antecedent	Hmotnost	<65;70)
	Paš	<65;70], <70;75)
Succedent	PQ	(<= <130;140))

Contingency table

	Succedent	NOT Succedent	
Antecedent	16	9	25
NOT Antecedent	222	745	967
	238	754	992

Values from contingency table:

a	16	16	a-frequency from the contingency table
b	9	9	b-frequency from the contingency table
c	222	222	c-frequency from the contingency table
d	745	745	d-frequency from the contingency table
r	25	25	r-frequency (a+b) from the contingency table
n	992	992	n-frequency (a+b+c+d) from the contingency table
Conf	0.64	0.64	Confidence (validity): $a/(a+b)$
DConf	0.06	0.0647773279	D-Confidence: $a/(a+b+c)$
EConf	0.77	0.7671370968	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.02	0.0161290323	Support: $a/(a+b+c+d)$
Cmplt	0.07	0.0672268908	Completeness: $a/(a+c)$
AvgDf	1.67	1.6675630252	Average difference: $a(a+b+c+d)/((a+b)(a+c))-1$

2) Kladná odchylka od průměru

- *Antecedent 0-3 atributů* : Atribut Boky, Hmotnost, Pas, Vyska – intervaly délky 1-3
 - *Succedent* : Atribut Frekvence - řez délky 1-2
 - $P \geq 1.0$ a $BASE \geq 20$
 - *Výsledek* : Nalezen 1 vztah
-

2) Kladná odchylka od průměru

Antecedent	Hmotnost	(<90;95)...<100;105))
	Pas	(<100;105)...<110;115))
	Vyska	(<170;175)...<180;185))
Succedent	Frekvence	pod 50, <50;60)

Contingency table

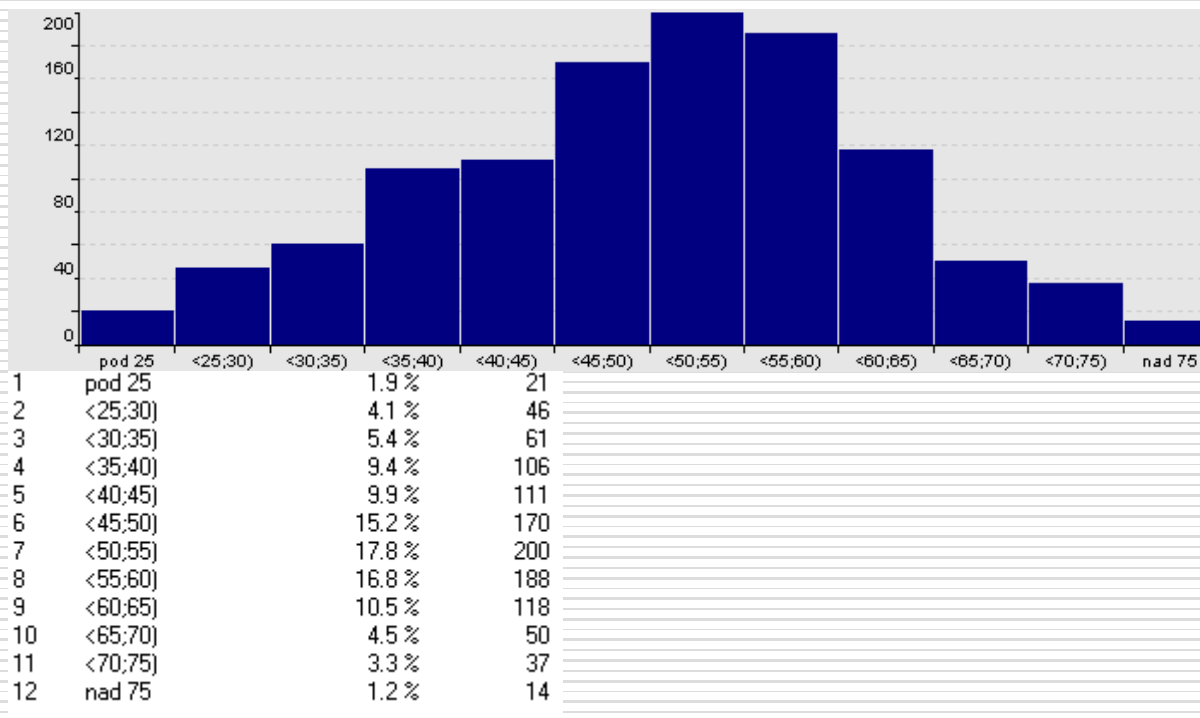
	Succedent	NOT Succedent	
Antecedent	20	39	59
NOT Antecedent	161	853	1014
	181	892	1073

Values from contingency table:

a	20	20	a-frequency from the contingency table
b	39	39	b-frequency from the contingency table
c	161	161	c-frequency from the contingency table
d	853	853	d-frequency from the contingency table
r	59	59	r-frequency (a+b) from the contingency table
n	1073	1073	n-frequency (a+b+c+d) from the contingency table
Conf	0.34	0.3389830508	Confidence (validity): $a/(a+b)$
DConf	0.09	0.0909090909	D-Confidence: $a/(a+b+c)$
EConf	0.81	0.8136067102	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.02	0.018639329	Support: $a/(a+b+c+d)$
Cmplt	0.11	0.1104972376	Completeness: $a/(a+c)$
AvgDf	1.01	1.0095514561	Average difference: $a(a+b+c+d)/((a+b)(a+c)) - 1$

Vliv dodatečné podmínky: Věk

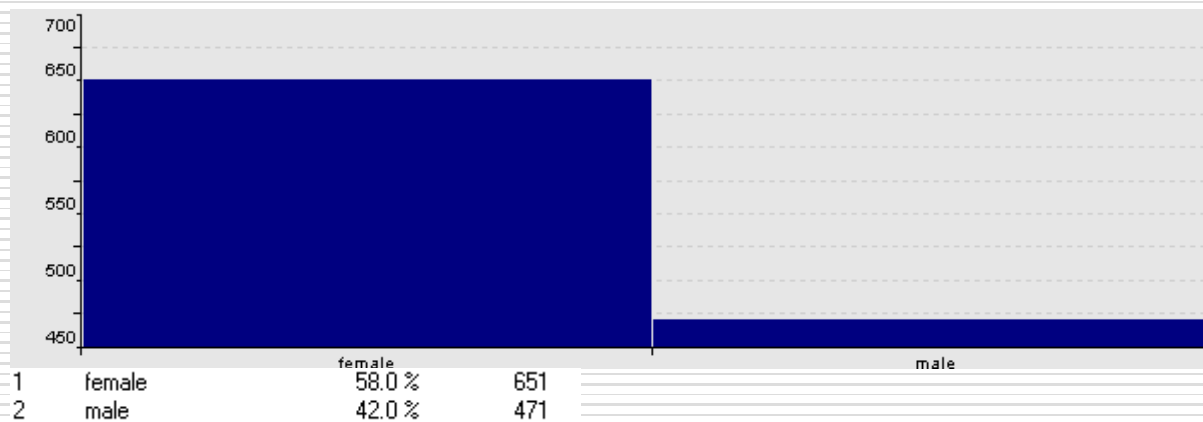
Atribut byl vytvořen ze sloupce „Vek“ pomocí rozdělení *Equidistant intervals* s délkou intervalu 5let .
•počet kategorií : 12



Vliv dodatečné podmínky: Pohlaví

Atribut vytvořen ze sloupce „Pohl“ pomocí *Each value – one category*.

•počet kategorií : 2



Vliv dodatečné podmínky

- Byla vytvořena podmínka tvaru :
 - Atribut Vek - Řez délky 1-4
 - Atribut Pohlaví - One Category - Female
 - Pro zkoumání vlivu podmínky byly vybrány vztahy tvaru $\varphi \Rightarrow p; \text{Base } \psi$. Tyto vztahy podmínku nespĺňují, zato byly nalezeny nové vztahy, které této podmínce vyhovovaly.

 - *Antecedent 0-3 atributů* : Atribut Boky, Hmotnost, Pas, Vyska - intervaly délky 1-3
 - *Succedent* : Atribut QRS - Řez délky 1 - 4
 - $P \geq 0.95$ a $\text{BASE} \geq 20$
 - *Podmínka* : Atribut Vek - Řez délky 1 - 4 & Atribut Pohlaví - One Category - Female
 - *Výsledek* : Nalezen 1 vztah
-

Antecedent	Hmotnost	<50;55], <55;60)
	Pas	(<= <65;70))
	Vyska	(<155;160)...<165;170))
Succedent	QRS	(<= <90;100))
Condition	Vek	(<= <35;40))
	Pohlaví	female

Contingency table

	Succedent	NOT Succedent	
Antecedent	20	1	21
NOT Antecedent	59	29	88
	79	30	109

Values from contingency table:

a	20	20	a-frequency from the contingency table
b	1	1	b-frequency from the contingency table
c	59	59	c-frequency from the contingency table
d	29	29	d-frequency from the contingency table
r	21	21	r-frequency (a+b) from the contingency table
n	109	109	n-frequency (a+b+c+d) from the contingency table
Conf	0.95	0.9523809524	Confidence (validity): $a/(a+b)$
DConf	0.25	0.25	D-Confidence: $a/(a+b+c)$
EConf	0.45	0.4495412844	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.18	0.1834862385	Support: $a/(a+b+c+d)$
Cmplt	0.25	0.253164557	Completeness: $a/(a+c)$
AvgDf	0.31	0.3140446052	Average difference: $a(a+b+c+d)/((a+b)(a+c))-1$

-
- *Antecedent 0-3 atributů* : Atribut Boky, Hmotnost, Pas, Vyska – intervaly délky 1-3
 - *Succedent* : Atribut PQ - Řez délky 1 – 7
 - $P \geq 0.95$ a $BASE \geq 25$
 - *Podmínka* : Atribut Vek - Řez délky 1 – 4 & Atribut Pohlaví – One Category - Female
 - Výsledek : Nalezen 1 vztah
-

Antecedent	Hmotnost	<60;65], <65;70]
	Pas	(<70;75]...<80;85))
Succedent	PQ	(<= <150;160))
Condition	Vek	(<= <35;40))
	Pohlaví	female

Contingency table

	Succedent	NOT Succedent	
Antecedent	27	1	28
NOT Antecedent	65	18	83
	92	19	111

Values from contingency table:

a	27	27	a-frequency from the contingency table
b	1	1	b-frequency from the contingency table
c	65	65	c-frequency from the contingency table
d	18	18	d-frequency from the contingency table
r	28	28	r-frequency (a+b) from the contingency table
n	111	111	n-frequency (a+b+c+d) from the contingency table
Conf	0.96	0.9642857143	Confidence (validity): $a/(a+b)$
DConf	0.29	0.2903225806	D-Confidence: $a/(a+b+c)$
EConf	0.41	0.4054054054	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.24	0.2432432432	Support: $a/(a+b+c+d)$
Cmplt	0.29	0.2934782609	Completeness: $a/(a+c)$
AvgDf	0.16	0.163431677	Average difference: $a(a+b+c+d)/((a+b)(a+c)) - 1$

-
- *Antecedent 0-2 atributů* : Atribut Boky, Hmotnost, Pas, Vyska – intervaly délky 1-3
 - *Succedent* : Atribut PQ - Řez délky 1 – 7
 - $P \geq 0.9$ a $BASE \geq 20$
 - *Podmínka* : Atribut Vek - Řez délky 1 – 4 & Atribut Pohlaví – One Category - Female
 - Výsledek : Nalezeno 17 vztahů, nejzajímavější je uveden je předchozí zobrazený vztah, druhý nejzajímavější z hlediska hodnoty p je vztah zobrazený níže
-

Antecedent	Hmotnost	<60;65], <65;70)
	Pas	<70;75], <75;80)
Succedent	PQ	(<= <150;160))
Condition	Vek	(<= <35;40))
	Pohlaví	female

Contingency table

	Succedent	NOT Succedent	
Antecedent	24	1	25
NOT Antecedent	68	18	86
	92	19	111

Values from contingency table:

a	24	24	a-frequency from the contingency table
b	1	1	b-frequency from the contingency table
c	68	68	c-frequency from the contingency table
d	18	18	d-frequency from the contingency table
r	25	25	r-frequency (a+b) from the contingency table
n	111	111	n-frequency (a+b+c+d) from the contingency table
Conf	0.96	0.96	Confidence (validity): $a/(a+b)$
DConf	0.26	0.2580645161	D-Confidence: $a/(a+b+c)$
EConf	0.38	0.3783783784	E-Confidence: $(a+d)/(a+b+c+d)$
Supp	0.22	0.2162162162	Support: $a/(a+b+c+d)$
Cmplt	0.26	0.2608695652	Completeness: $a/(a+c)$
AvgDf	0.16	0.1582608696	Average difference: $a(a+b+c+d)/[(a+b)(a+c)]-1$