COUNTING EVENTS **C2**

Questions	Answers
How can we count possibilities?	By using the numbers in Pascal's triangle
How can we predict unpredictable numbers?	We 'post-dict' that the average number is 8.2 with the deviation 2.3. We 'pre-dict' that the next number, with 95% probability, will fall in the confidence interval 8.2 ± 4.6 (average ± 2*deviation)

1 COUNTING UNSYSTEMATIC EVENTS

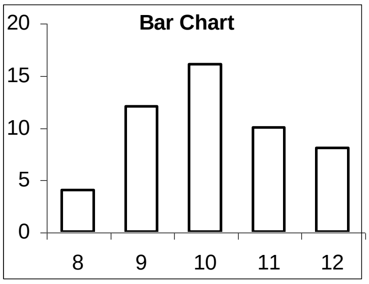
Question1. How to count events occurring randomly?
Answer. By working out statistics for the events, their frequencies and their average.
Exemple1. To the question 'How old are you?' the answers are arranged in a table showing the frequencies of the answers. From the distribution the average answer and the average deviation is calculated. From this 'post-diction' a pre-diction can be made saying that the following answer, with 95% probability, will fall in the 'confidence interval' I = average answer ± 2*average deviation.

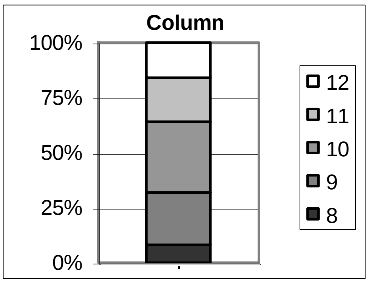
Observations x	Frequency f	Relative frequency p	Cumulative frequency	Total years lived	Deviation from the mean	Average deviation
8	4	4/50 = 0,08 = 8%	8%	8*4 = 32	10.1 - 8 = 2,1	2.1^2*4 = 17.64
9	12	12/50 = 0,24 = 24%	32%	9*12 = 108	10.1 - 9 = 1.1	1.1^2*12 = 14.52
10	16	16/50 = 0,32 = 32%	64%	10*16 = 160	10.1 - 10 = 0.1	0.1^2*16 = 0.16
11	10	10/50 = 0,20 = 20%	84%	11*10 = 110	11 - 10.1 = 0.9	0.9^2*10 = 8.1
12	8	8/50 = 0,16 = 16%	100%	12*8 = 96	12 - 10.1 = 1.9	1.9^2*8 = 28.88
Total	50	1,00 = 100%	-	506	-	69.3
Average, mean:				506/50 = 10.1		(√69.3)/50 = 1.2

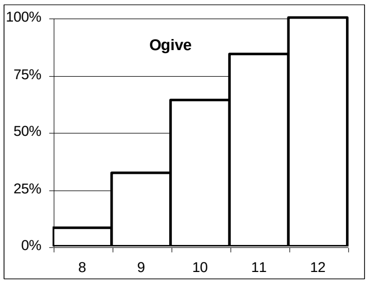
The frequency-table can be illustrated graphically both in case of grouped and non-grouped observations:

OBS	FRQ
8	4
9	12
10	16
11	10
12	8

DoubleClick &Edit



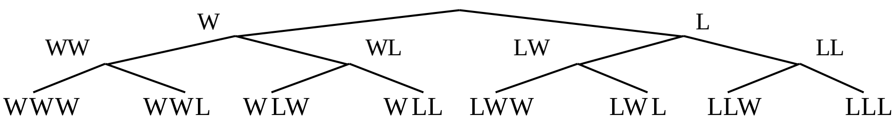




Exercise. Observe the sum of 2 dices 10 times. Set up a frequency table and calculate the confidence interval. Repeat the exercise 20 times, and 30 times. Repeat the exercise with the theoretical frequencies p(x=2)= 1/36, p(x=3) = 2/36, p(x=4) = 3/36, … , p(x=12) = 1/36. x is called a random variable or unpredictable variable.

2 COUNTING SYSTEMATIC EVENTS

Question1. How to count events (pools) in 3 repetitions of a 2-option Win/Loose-game?
Answer. By repeated multiplying.



game	pools	
1	2:	W, L
2	2*2:	WW, WL, LW, LL
3	2*2*2:	WWW,WWL,…

Key to 3 2-option games: 2 options for game 1; each having 2 options for game 2 totalling 2*2 events; each having 2 options for game 3 totalling (2*2)*2 = 2^3 events.
Prediction. There are r^n different events in n repetitions of an r-option game.
Exercise. Check the prediction with 4 2-option games, 4 3-option games, 4 4-option games, etc.
Question2. How to count letter plates?
Answer. By counting ordered arrangements, permutations.

Letters:	1 (A)	2 (A,B)	3 (A,B,C)	4 (A,B,C,D)	5 (A,B,C,D,E)
Plates:	1	2	6	?	?
	A	AB	ABC		
		BA	ACB		
			BAC		
			BCA		
			CAB		
			CBA		

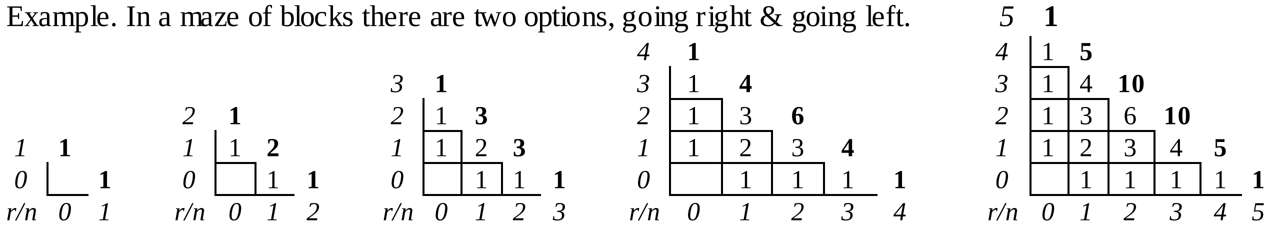
Key to 3-letter plates: 3 options for letter 1, 2 for letter 2, 1 for letter 3 totalling 3*2*1 = 3! plates.
Prediction. There are n! different n-letter plates (n! is read n factorial).
Exercise. Check the prediction with 4-letter plates and 5-letter plates, etc.
Question3. How to count restricted letter plates only consisting of 2 letters of 5.
Answer. By counting restricted permutations.

Letters:	1of2 (A,B)	2of3 (A,B,C)	2of4 (A,B,C,D)	2of5
Plates:	2	6	?	?
	A	AB		
	B	AC		
		BC		
		BA		
		CA		
		CB		

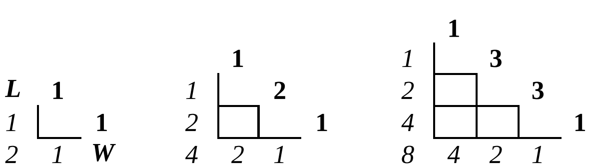
Key to 2of5 letter plates: 5 options for letter 1, 4options for letter 2 totalling 5*4=5*4*(3!/3!)=5!/3! Plates.
Prediction. There are n!/(n-r)! different n-of-r letter plates.
Exercise. Check the prediction with 2of4 and 3of5letter plates.
Question4. How to count lotto tickets in a lotto game guessing 2of5 letters.
Answer. By using the binomial numbers B(n,r) to count unordered arrangements, combinations.

Letters:	1of2 (A,B)	2of3 (A,B,C)	2of4 (A,B,C,D)	2of5
Plates:	2	6/2	12/2	?
	A	AB	AB	
	B	AC	AC	
		BC	AD	
		~~BA~~	BC	
		~~CA~~	BD	
		~~CB~~	CD	

Key to 2of5 lotto tickets: 5 options for letter 1, 4options for letter 2 totalling 5*4=5*4*(3!/3!)=5!/3! tickets giving 2*the total number of tickets. Thus B(5,2) = 5!/3!/2! = 5!/(3!*2!) = 5!/[(5-2)!*2!]
Prediction. There are B(n,r) = n!/[(n-r)!*r!] different n-of-r lotto tickets.
Exercise. Check the prediction with 2of4 and 3of5letter lotto tickets.
Question5. Is there a pattern in the binomial numbers B(n,r)?
Answer. The binomial numbers B(n,r) form a triangle (Pascall's triangle).
Example. In a maze of blocks there are two options, going right & going left.

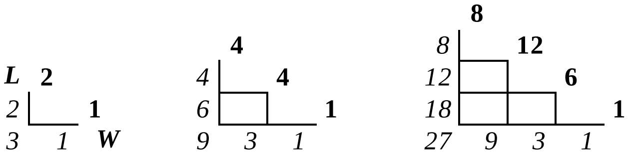


Prediction. B(n,0) = B(0,n) = n. B(n,r) = B(n-r,r). B(n,r) = B(n,r-1)+B(n-1,r). B(0,0) is defined as 1.
Exercise. Check the prediction with the numbers B(6,r) and B(8,r).
Question6. How to count the winners in 4 2-option Win/Loose-games?
Answer. By the binomial numbers B(n,r).



Key. Among 2 persons, 1 wins 0 times and 1 wins 1 time (1,1 persons win 0,1 times) in 1 2-option game.
Among 4 persons 1,2,1 persons win 0,1,2 times in 2 2-option games.
Among 8 persons 1,3,3,1 persons win 0,1,2,3 times in 3 2-option games.
Prediction1. Among 16 persons 1,4,6,4,1 persons win 0,1,2,3 times in 4 2-option games.
Or among 2^4 persons B(4,0), B(4,1), B(4,2), B(4,3), B(4,4) persons win 0,1,2,3 times in 4 2-option games.
Exercise. Check the binomial prediction with 4 2-option games.
Prediction2. Among 2^n persons B(n,r) persons win r times in n 2-option games.
Exercise. Check the binomial prediction with 6, 7 and 8 2-option games.

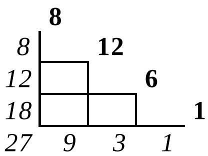
Question7. How to count the winners in 4 3-option games (1 Win and 2 Loose)?
Answer. By weighted binomial numbers.



Persons in 3 3-option games

1 * 2/3*2/3*2/3 * 27	B(3,3)*(1/3)^0*(2/3)^3 * 27
3 * 1/3*2/3*2/3 * 27	B(3,2)*(1/3)^1*(2/3)^2 * 27
3 * 1/3*1/3*2/3 * 27	B(3,1)*(1/3)^2*(2/3)^1 * 27
1 * 1/3*1/3*1/3 * 27	B(3,0)*(1/3)^3*(2/3)^0 * 27

Key. Among 3 persons 2 win 0 times and 1 wins 1 time (2,1 persons win 0,1 times) in 1 3-option game.
Among 9 persons 4,4,1 persons win 0,1,2 times in 2 3-option games.
Among 27 persons 8,12,6,1 persons win 0,1,2,3 times in 3 3-option games.
Prediction1. Among 81 persons 16,32,24,8,1 persons win 0,1,2,3,4 times in 4 3-option games.
Or Among 3^4 persons B(4,0)*16, B(4,1)*8, B(4,2)*4, B(4,3)*2, B(4,4)*1 persons win 0,1,2,3,4 times.
Or Among 3^4 persons B(4,0)*2^4, B(4,1)*2^3, B(4,2)*2^2, B(4,3)*2^1, B(4,4)*2^0 persons win 0,1,2,3,4 times.
Exercise1. Check the binomial prediction with 4 3-option games.
Prediction2. Among 3^n persons B(n,r)*1^r*2^(n-r) persons win r times in n 3-option 1W/2L-games
Exercise2. Check the binomial prediction with 5, 6 and 7 3-option games, etc.
Prediction3. Among 4^n persons B(n,r)*1^r*3^(n-r) persons win r times in n 4-option 1W/3L-games
Exercise3. Check the binomial prediction with 2, 3 and 4 4-option games, etc.
Prediction4. Among k^n persons B(n,r)*1^r*(k-1)^(n-r) persons win r times in n k-option 1W/(k-1)L-games.
Exercise4. Check the binomial prediction with 4 5-option games and 4 6-option games.
Prediction5. Among k^n persons B(n,r)*q^r*(k-q)^(n-r) persons win r times in n k-option qW/(k-q)L-games.
Exercise5. Check the binomial prediction with 3 5-option games and a 4 6-option games both with 2 wins.
Question8. How to predict the chance for winning?
Answer. By using probabilities.
Example1. A dice is manipulated changing the 6 and 5 to 1 and the 4 to 2. Thus out of 6 outcomes 1 occurs 3 times (3 = 3/6*6), 2 occurs 2 times (2 = 2/6*6) and 3 occurs 1 time (1 = 1/6*6). The number 3/6 is called the chance or probability for the event 1: p(x=1) = 3/6. Likewise p(x=2) = 2/6, p(x=3) = 1/6.
Example2. In a 3-option game the chance to win is p = 1/3 if the outcomes are equally likely. Thus the probabilities can be found from the binomial numbers:



Persons		*Probabilities*
1 * 2/3*2/3*2/3 * 27	B(3,3)*(1/3)^0*(2/3)^3 * 27	p(x=0) = B(3,3)*(1/3)^0*(2/3)^3
3 * 1/3*2/3*2/3 * 27	B(3,2)*(1/3)^1*(2/3)^2 * 27	p(x=1) = B(3,2)*(1/3)^1*(2/3)^2
3 * 1/3*1/3*2/3 * 27	B(3,1)*(1/3)^2*(2/3)^1 * 27	p(x=2) = B(3,1)*(1/3)^2*(2/3)^1
1 * 1/3*1/3*1/3 * 27	B(3,0)*(1/3)^3*(2/3)^0 * 27	p(x=3) = B(3,0)*(1/3)^3*(2/3)^0

The binomial probabilities p(x = r) = B(n,r)* p^r*(1-p)^(n-r), where x counts the number of wins, are cumulated before being tabulated, e.g. in case of 5 3-option games with a winning chance p = 1/3:

x	p	∑p	P(x >= 3) at least 3W	P(x <= 4) at most 4W	P(x = 2) Precisely 2W	P(1 <= x <= 3) Between 1W & 3W
0	0.132	0.132				-0.132
1	0.329	0.461			-0.461	
2	0.329	0.790	-0.790		+0.790	
3	0.165	0.955				+0.955
4	0.041	0.996		+0.996		
5	0.004	1.000	+1.000			
			0.210	**0.996**	**0.329**	**0.832**

Thus there is a 21% chance for winning at least 3 times of 5, etc.
Example2. The average wins in a n=5 game depends on the winning chance (here p=1/3, p=1/4 and p=1/2):

x	p	x*p		x	p	x*p		x	p	x*p
0	0.132	0		0	0.237	0		0	0.031	0
1	0.329	0.329		1	0.396	0.396		1	0.156	0.156
2	0.329	0.658		2	0.264	0.527		2	0.313	0.625
3	0.165	0.495		3	0.088	0.264		3	0.313	0.938
4	0.041	0.164		4	0.015	0.059		4	0.156	0.625
5	0.004	0.020		5	0.001	0.005		5	0.031	0.156
		m=1.666				**m=1.25**				**m=2.5**

Prediction. The average wins in n games with the winning chance p is m = n*p
Exercise. Check the prediction with n=3 and n=4 games with the winning chance p=1/3, p=3/4, p=½, p=2/5.

Example3. The average deviation from the average depends on the winning chance (here p=1/4 and p=1/2):

x	p	x*p	c = x-m	c^2*p		x	p	x*p	c = x-m	c^2*p
0	0.237	0	-1.25	0.371		0	0.031	0	-2.5	0.195
1	0.396	0.396	-0.25	0.025		1	0.156	0.156	-1.5	0.352
2	0.264	0.527	0.75	0.148		2	0.313	0.625	-0.5	0.078
3	0.088	0.264	1.75	0.269		3	0.313	0.938	0.5	0.078
4	0.015	0.059	2.75	0.111		4	0.156	0.625	1.5	0.352
5	0.001	0.005	3.75	0.014		5	0.031	0.156	2.5	0.195
	*n*p* =	**m=1.25**		v = 0.938			*n*p* =	**m=2.5**		v = 1.25
			√*n*p* =	√v= 0.968					√*n*p* =	√v= 1.118

Prediction. The average deviation D in n games with the winning chance p is D = √(n*p*(1-p)) = √n*p, where the average of p and 1-p is p = √(p*(1-p))

Exercise. Check the prediction with n=3 and n=4 games with the winning chance p=1/3, p=3/4, p=½, p=2/5.

Remark. Dividing the binomial numbers B(n,r)*q^r*(k-q)^(n-r) with k^n gives the binomial probabilities:

$$\frac{B(7,r) * 1\text{^}r * 3\text{^}(7-r)}{4\text{^}7} = B(7,r) * \frac{1\text{^}r * 3\text{^}(7-r)}{4\text{^}r * 4\text{^}(7-r)} = B(7,r) * \frac{1\text{^}r}{4\text{^}r} * \frac{3\text{^}(7-r)}{4\text{^}(7-r)} = B(7,r) * \frac{1}{4}\text{^}r * \frac{3}{4}\text{^}(7-r) = B(7,r) * p\text{^}r * (1-p)\text{^}(7-r)$$

Question9. How to predict the unpredictable when events are systematic?

Answer. The number of wins can be predicted as x' = x average ± 2*average deviation = n*p ± 2*√n*p.
And the percentage of wins can be predicted as p' = x'/n = (n*p ± 2*√n*p)/n = p ± 2*p/√n.

Example1. In a n=7 game with the winning chance p=2/5 the numbers of wins is predicted to be 7*2/5 ± 2*√7*√(2/5*3/5) = 2.8 ± 2.6 = [0.2;5.4]. So winning 0, 6 or 7 times is less than 5% probable.

Exercise. What is the prediction for the sum when throwing 2 dices? And when throwing 3 dices? Check it.

Question10. How does the mean of a sample vary?

Answer. From a population with a mean m and a deviation D the experiment 'examine a member' is repeated n times. The mean of this sample is an unpredictable event having m as mean and D/√n as deviation. Its distribution is called a normal distribution which is tabulated as cumulated frequencies, e.g. Φ(1.220) = 0.889.

Example. Throwing a dice has 6 events with the mean 3,5 and the deviation 1,71. A sample is produced by throwing a dice 10 times (or throwing 10 dices once). The sample-mean is an unpredictable variable with the mean 3,5 and the deviation 1,71/√10.

Exercise. Check the prediction by 12 times throwing ten dices each time observing the mean.

Question11. Are there any short cuts to the binomial distribution?

Answer. A binomial distribution can be approximated by a normal distribution if n*p>5 and n*(1-p)>5.
If n=20 and p=0.3 then m=n*p=20*0.3=6 (and n*(1-p)=20*0.7=14) and D =√n*p=√20*√(0.3*0.7)=2.05

Factual number	Approximated number
P(x<=8) = 0.887	P(x<=8) = Φ((8+0.5-6)/2.05) = Φ(1.220) = 0.889

Exercise. Check several examples of the normal approximation prediction of binomial probabilities.

Question12. How to win in a 2person game?

Answer. By mixing the strategies unpredictably.

Example. The players A and B can choose between two strategies, Paper and Stone. The gain paid from B to A depends on the combination (A,B): g(P,P) = -1, g(S,S) = -1, g(P,S) = 1, g(S,P) = 2. A hopes for the gain 2 and chooses S; B sees this and chooses S; A sees this and chooses P; B sees this and chooses P, etc.

A broker proposes the following solution: both A and B choose a mixed strategy P/S = x/(1-x) and P/S = y/(1-y). The gain then is g = -1*x*y + 1*x*(1-y) + 2*(1-x)*y - 1*(1-x)*(1-y) = -1 - 5*x*y + 2*x + 3*y.
y = 0 lets g = -1+2x; y = 1 lets g = 2-3x. The intersection point is x = 3/5 = 60% and g = 1/5.
x = 0 lets g = -1+3y; x = 1 lets g = 1-2y. The intersection point is y = 2/5 = 40% and g = 1/5.
Thus the fair result of this game is that B pays to A 1/5 per game. Else A weighs P 60%, B weighs P 40%.

Exercise. Check the prediction by playing the game with the weighed strategies determined by 10 cards.

3 REVERSING STATISTICS

Question1. How to do reversed calculation in statistics?

Answer. By testing hypothesis.

Example1. There is one red card, or is there?

I assume I got what I ordered, 1 red and 1 black card. To check I randomly draw 1 card 3 times, giving 0 red cards. The probability for this event is p(x=0) = B(n,r)* p^r*(1-p)^(n-r) = B(3,0)* 1/2^0*(1/2)^3 = 1/8.
To reject my hypothesis 1R1B the probability must be less than 5%, so I have to draw n=5 times since p(x=0) = B(n,r)* p^r*(1-p)^(n-r) = B(n,0)* 1/2^0*(1/2)^n = ½^n = 5%, gives n = log5%/log½ = 4.3.

Example2. There is only one red card, or is there?

I assume I got what I ordered, 1 red & 2 black cards. To check I randomly draw 1 card 10 times, giving 6 red

cards. The probability for the event x>=6 with n,p = 10,1/3 is p(x>=6) = 1-p(<=5) = 1-0.923 = 0.077.
I cannot reject my hypothesis 1R2B since the probability is higher than 5%.
Next time I got 7 red cards. The probability for the event x>=7 in an n,p = 10,1/3 game is p(x>=7) = 1-p(<=6) = 1-0.980 = 0.02. Now I reject my hypothesis since the probability is less than 5%.
Example 3. I got 20 cards, how many are red? To check I randomly draw 1 card 10 times, giving 6 red cards. The following hypothesis can be rejected: The number of red cards is 0, 1, 2, …, 6, 18, 19, 20.
With n,p = 10 , 1/20 the probability for the event x>=6 is p(x>=6) = 1-p(<=5) = 1-1.000 = 0
With n,p = 10 , 6/20 the probability for the event x>=6 is p(x>=6) = 1-p(<=5) = 1-0.953 = 0.047
With n,p = 10 , 17/20 the probability for the event x<=6 is p(x<=6) = 0.050
With n,p = 10 , 19/20 the probability for the event x<=6 is p(x<=6) = 0.001
This result can be predicted by the confidence interval I = p ± 2*$\underline{p}$/√n = 6/10 ± 2*$\underline{0.6}$/√10 = 0.60 ± 0.31 = [0.29;0.91]. Thus the hypotheses that can be rejected are from 0/20 to 6/20 and from 18/20 to 20/20.
Exercise. Take 2 like bags. Put 2 reds & 1 black ball in one and 1 red & 2 blacks in the other. Choose 1 bag. Repeat n times the experiment 'take out 1 ball' until you can form a hypothesis. How sure can you be?

4 COMPARING REPRESENTATIONS

Question1. How to compare parts?
Answer. Through conditional percentages.
Example 1. We compare two decks of cards with and without red Kings (I and II).

I	Spades	NonSpades	Total
CourtCards	3	9	12
NumberCards	10	30	40
Total	13	39	52

II	Spades	NonSpades	Total
CourtCards	3	7	10
NumberCards	10	30	40
Total	13	37	50

I. The unconditional probability or percentage of CoutCards (among all) = p(C I all) = p(C) = 12/52 = 23.1%.
The conditional probability or percentage of CoutCards (among the spades) = p(C|S) = 3/13 = 23.1%.
Since p(C|Spade)=p(C) we say that 'among the spades the CourtCards are neither over- or under-represented.'
The conditional probability or percentage can be shown in a contingency table or pivot table:

I after colour	Spades	NonSpades	Total
CourtCards	23.1%	23.1%	23.1%
NumberCards	76.9%	76.9%	76.9%
Total	100%	100%	100%

I after kind	Spades	NonSpades	Total
CourtCards	25%	75%	100%
NumberCards	25%	75%	100%
Total	25%	75%	100%

II. The unconditional probability or percentage of CoutCards (among all) = p(C I all) = p(C) = 10/50 = 20%.
The conditional probability or percentage of CoutCards (among the spades) = p(C|S) = 3/13 = 23.1%.
Since p(C I Spade) > p(C) we say that 'among the spades the CourtCards are over-represented.'
Since p(C I NonSpade) < p(C) we say that 'among the non-spades the CourtCards are under-represented.'
The conditional probability or percentage can be shown in a contingency table or pivot table:

II acc. colour	Spades	NonSpades	Total
CourtCards	23.1%	18.9%	20%
NumberCards	76.9%	81.1%	80%
Total	100%	100%	100%

II acc. kind	Spades	NonSpades	Total
CourtCards	30%	70%	100%
NumberCards	25%	75%	100%
Total	26%	74%	100%

Example 2. A big population is distributed according to gender and smoking habit. To get an estimate of the percentages we draw a sample from a population creating an uncertainty to the percentages D=2*$\underline{p}$/√n.
Thus p(Girl I Smoker) = 27.3% and n = 550 gives D = 2*√(27.3%*72.7%)/√550 = 3.8%.
We set up frequency tables and contingency tables:

Sample	Smoker	NonSmoker	Total	*D*
Female	150	270	420	*4,7%*
Male	400	500	900	*3,3%*
Total	550	770	1320	*2,7%*
D	*3,8%*	*3,4%*	*2,6%*	

DoubleClick &Edit

Acc. Gender	Smoker	NonSmoker	Total
Female	27,3%	35,1%	31,8%
Male	72,7%	64,9%	68,2%
Total	100,0%	100,0%	100,0%

Acc. Smoking	Smoker	NonSmoker	Total
Female	35,7%	64,3%	100,0%
Male	44,4%	55,6%	100,0%
Total	41,7%	58,3%	100,0%

A deviation is significant if it exceeds the uncertainty. From the distribution acc. to gender table we can say:
Among the smokers the females are significantly under-represented (31.8%-27.3% = 4.5% > 3.8%).
Among the non-smokers the females are not significantly over-represented (35.1%-31.8% = 3.3% < 3.4%).
In the population the females are significantly under-represented (50%-31.8% = 18.2% > 2.6%).
Exercise. What can be said about the males? What can be said from the distribution after smoking table?