ΤΕΙ ΚΑΒΑΛΑΣ-ΣΤΕΦ
ΤΜΗΜΑ ΒΙΟΜΗΧΑΝΙΚΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
Τυπικό Εξάμηνο Σπουδών: 4ο

ΜΑΘΗΜΑ: ΔΙΑΧΕΙΡΙΣΗ ΠΟΙΟΤΗΤΑΣ (Ε)

2Η ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ

201. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΣΤΑΤΙΣΤΙΚΗΣ – ΔΕΙΓΜΑΤΙΚΑ ΜΕΤΡΑ
202. ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

ΒΑΣΙΚΕΣ ΈΝΝΟΙΕΣ ΣΤΑΤΙΣΤΙΚΗΣ
ΔΕΙΓΜΑΤΙΚΑ ΜΕΤΡΑ

Γενικά

Τα αριθμητικά μέτρα σύνοψης διακρίνονται σε μέτρα θέσης (όπως η μέση τιμή και η διάμεσος), μέτρα διασποράς (όπως η τυπική απόκλιση και το εύρος), μέτρα ασυμμετρίας, κυρτότητας.

Μέτρα Θέσης

- Μέση Τιμή
- Διάμεσος

Μέτρα διασποράς

- Τυπική Απόκλιση
- Εύρος
- Διακύμανση

Μέτρα Ασυμμετρίας και Μέτρα Κυρτότητας

Άσκηση 201η

Έστω λοιπόν ότι μια βιομηχανία κατασκευάζει ένα εξάρτημα που χρησιμοποιείται στις μητρικές πλακέτες (motherboards) των υπολογιστών, και κατασκευάζεται από δύο γραμμές παραγωγής (Α και Β) της βιομηχανίας.

Πολύ βασικό σημείο για την λειτουργικότητα του εξαρτήματος αποτελεί η ποιότητα της επιφάνειας του, η οποία αξιολογείται με μετρήσεις τραχύτητας.

Για τον σκοπό αυτό, από κάθε γραμμή παραγωγής επιλέγονται **τυχαία 40** δείγματα σε δεδομένη επιφάνεια των οποίων εκτελείται δοκιμή τραχυμέτρησης, όπου υπολογίζεται η τραχύτητας (Ra) σε μm.

Τα αποτελέσματα των δύο σειρών μετρήσεων έχουν ως εξής:

Γραμμή Παραγωγής Α

30,5	31,2	33,1	30,7	32,5	31,5	31,8	32,0
32,5	29,4	29,8	29,1	29,0	30,5	30,2	28,3
30,5	31,2	31,8	28,3	30,5	31,2	34,5	27,8
28,7	30,6	27,6	27,5	30,4	32,5	29,2	30,4
33,6	31,5	32,0	31,2	32,0	32,5	34,6	27,9

Γραμμή Παραγωγής Β

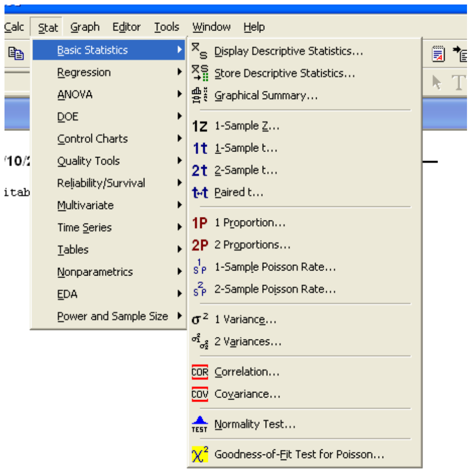
37,0	37,4	37,2	36,9	37,4	36,5	36,6	37,9
37,1	36,5	36,0	37,4	36,2	36,4	35,8	35,4
32,5	38,6	34	37,5	37,4	36,5	35,1	35,2
37,4	37,2	36,8	36,9	35,4	31,8	33,1	37,6
35,3	35,6	37,6	33,5	35,7	32,4	36,2	37,4

Για τα **40** δείγματα της ***Α Γραμμής Παραγωγής*** και με την χρήση του Minitab 15 να υπολογίσετε τα στατιστικά μέτρα της μεταβλητής της τραχύτητας του δείγματος.

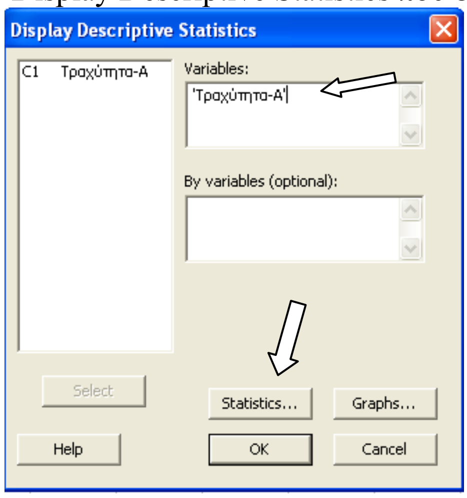
Απάντηση

Για να πάρουμε τα περιγραφικά στατιστικά της μεταβλητής C1-Τραχύτητα εργαζόμαστε ως εξής:

1. Από τη γραμμή μενού επιλέγουμε **Stat → Basic Statistics → Display Descriptive Statistics.**



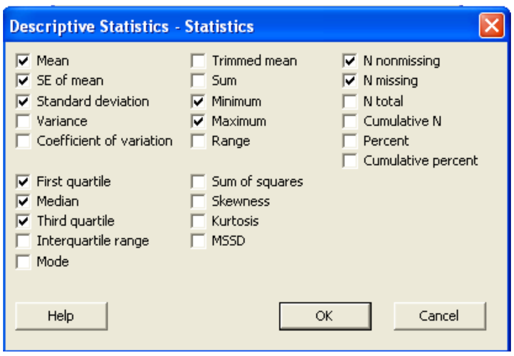
2. Στο πλαίσιο διαλόγου Display Descriptive Statistics που εμφανίζεται:



(α) Στο πλαίσιο Variables επιλέγουμε τη μεταβλητή **C1 Τραχύτητα-Α** από τον αριστερό κατάλογο των διαθέσιμων στηλών,

(β) Επιλέγουμε στο κουμπί **Statistics.**

3. Στο πλαίσιο διαλόγου Descriptive Statistics - Statistics που εμφανίζεται επιλέγουμε όλα τα δειγματικά στατιστικά μέτρα που επιθυμούμε.



4. Επιλέγουμε δύο φορές OK και στο Session Window έχουμε τα εξής αποτελέσματα:

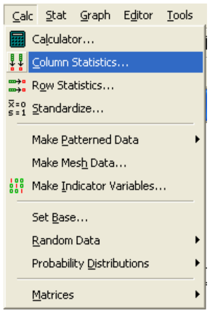
```
Descriptive Statistics: Τραχύτητα-A

             Total
Variable     Count   N  N*  CumN  Percent  CumPct    Mean  SE Mean  TrMean
Τραχύτητα-A     40  40   0    40      100     100  30,753    0,287  30,719

                                                      Sum of
Variable     StDev  Variance  CoefVar       Sum     Squares  Minimum      Q1
Τραχύτητα-A  1,818     3,304     5,91  1230,100   37957,510   27,500  29,250

                                                                      N for
Variable     Median      Q3  Maximum  Range    IQR               Mode  Mode
Τραχύτητα-A  30,650  32,000   34,600  7,100  2,750  30,5; 31,2; 32,5      4


Variable     Skewness  Kurtosis   MSSD
Τραχύτητα-A      0,03     -0,42  3,295
```

Τα στατιστικά μέτρα που μπορεί να εμφανίσει το MINITAB 15 είναι συνοπτικά τα εξής:

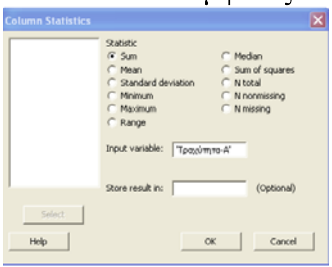
Στατιστικό Μέτρο	**Σύμβολο MINITAB**	**Τι υπολογίζει:**
Number of nonmissing values	N	Τον αριθμό των κελιών που περιλαμβάνουν τιμές.
Number of missing values	N*	Τον αριθμό των κελιών που δεν περιλαμβάνουν τιμές (*).
Total Number	Total Count	Το άθροισμα των δύο παραπάνω.
Cumulative number	CumN	Την αθροιστική συχνότητα των παρατηρήσεων-τιμών για κάθε μεταβλητή ομαδοποίησης. Όταν υπάρχει μια στήλη ταυτίζεται με το N.
Percent	Percent	Το ποσοστό μιας κατηγορίας στο σύνολο. Όταν υπάρχει μια στήλη υπολογίζει το N/Total Count.
Cumulative percent	CumPer	Την αθροιστική κατανομή μιας κατηγορίας. Όταν υπάρχει μια στήλη ταυτίζεται με την Percent.
Mean	Mean	Τον αριθμητικό Μέσο Όρο
Trimmed mean	TrMean	Τον αριθμητικό Μέσο Όρο που προκύπτει αν αφαιρέσουμε ένα ποσοστό παρατηρήσεων και συγκεκριμένα το 5% των μεγαλύτερων και μικρότερων αριθμητικά παρατηρήσεων.
Standard error of mean	SE Mean	Την ακρίβεια με την οποία η δειγματική μέση τιμή προσεγγίζει την μέση τιμή (μ) του πληθυσμού, δηλαδή το τυπικό σφάλμα του μέσου όρου.
Standard deviation	StDev	Την δειγματική Τυπική Απόκλιση
Variance	Variance	Την δειγματική Διακύμανση
Coefficient of variation	CoefVar	Τον Συντελεστή Μεταβλητότητας του συντελεστή διακύμανσης
Sum	Sum	Το άθροισμα των τιμών
Minimum	Minimum	Την μικρότερη τιμή
Maximum	Maximum	Την μεγαλύτερη τιμή
Range	Range	Το δειγματικό εύρος
Median	Median	Την Διάμεσο του δείγματος
First and third quartiles	Q1, Q3	Το πρώτο και τρίτο Τεταρτημόριο του δείγματος
Interquartile range	IQR	Το ενδοτεταρτημοριακό εύρος του δείγματος
Mode		Την τιμή που εμφανίζεται με την μεγαλύτερη συχνότητα στο δείγμα
Sums of squares	Sums of squares	Το άθροισμα τετραγώνων των τιμών

Skewness	Skewness	Τον συντελεστή λοξότητας
Kurtosis	Kurtosis	Τον συντελεστή κύρτωσης
MSSD	MSSD	$MSSD = \dfrac{\sum (X_{i+1} - X_i)^2}{2(n-1)}$

Τα στατιστικά μιας στήλης μπορούν επίσης να υπολογιστούν και **μεμονωμένα** με την εντολή **Column Statistics**.
Από τη γραμμή μενού επιλέγουμε **Calc → Column Statistics.**



Στο πλαίσιο διαλόγου Column Statistics που εμφανίζεται



Από τη περιοχή Statistic επιλέγουμε το στατιστικό μέτρο το οποίο θέλουμε να υπολογίσουμε**,**
Στο πλαίσιο **Input variable**: επιλέγουμε **C1 Τραχύτητα-Α** από τον αριστερό κατάλογο και μετά **OK** και στο Session Window παίρνουμε τα αποτελέσματα.

Πρέπει να παρατηρήσουμε εδώ ότι στο **Column Statistics** μπορούμε να επιλέξουμε μόνο ένα στατιστικό μέτρο. Αν στην συνέχεια θέλουμε και κάποιο άλλο τότε από τη γραμμή εργαλείων πατάμε στο κουμπί **Edit**  **Last Dialog.**
Στο πλαίσιο διαλόγου **Column Statistics** που επανεμφανίζεται επιλέγουμε το επόμενο στατιστικό μέτρο που θέλουμε να υπολογίσουμε.

B) Τι συμπεράσματα προκύπτουν για την κατανομή των δεδομένων από τους Συντελεστές Λοξότητας και Κύρτωσης;
Απάντηση:
Από τα αποτελέσματα του MINITAB έχουμε:

```
Variable      Skewness  Kurtosis   MSSD
Τραχύτητα-Α       0,03     -0,42  3,295
```

Παρατηρούμε ότι Skewness=0,03 >0 συνεπώς η κατανομή των δεδομένων σύμφωνα με όσα αναφέρονται στην σελίδα 118-119 των σημειώσεων του Μαθήματος παρουσιάζει «θετική λοξότητα».
Επίσης Kurtosis= -0,42 <3 συνεπώς η κατανομή των δεδομένων είναι «πλατύκυρτη».

Εργασία

Να επαναληφθούν τα ζητήματα Α), Β) για την Γραμμή Παραγωγής Β.

ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

Γενικά

Ένα τυχαίο πείραμα στο οποίο υπάρχουν δυο μόνο δυνατά αποτελέσματα λέγεται δοκιμή Bernoulli. Το ένα από τα δυο αυτά αποτελέσματα ονομάζεται συνήθως αυθαίρετα «επιτυχία» και το άλλο «αποτυχία».

Έστω ότι μια δοκιμή Bernoulli με πιθανότητα επιτυχίας p επαναλαμβάνεται n φορές, έτσι ώστε το αποτέλεσμα μιας δοκιμής να είναι ανεξάρτητο από το αποτέλεσμα σε οποιαδήποτε άλλη δοκιμή. Τότε ο αριθμός X των επιτυχιών στο πείραμα αυτό είναι μια τυχαία μεταβλητή που μπορεί να πάρει τις τιμές 0,1,2,...,n.

Γενικά, η κατανομή πιθανοτήτων του αριθμού X των επιτυχιών σε n ανεξάρτητες δοκιμές Bernoulli με σταθερή πιθανότητα επιτυχίας p δίνεται από τη σχέση:

$$P(X = x) = \binom{n}{x} p^{x} (1-p)^{n-x}$$

Γενικά η διωνυμική κατανομή εφαρμόζεται σε πειράματα τύχης που χαρακτηρίζονται από τις ακόλουθες συνθήκες:

1. Υπάρχουν δυο μόνο λογικά αποτελέσματα στο πείραμα τύχης: να πραγματοποιηθεί, για παράδειγμα, το ενδεχόμενο A ή να μην πραγματοποιηθεί. Οπότε, αν οι αντίστοιχες πιθανότητες είναι p και q, τότε ισχύει ότι: p+q=1.
2. Το πείραμα τύχης εφαρμόζεται n ανεξάρτητες φορές ή εκτελούνται n ανεξάρτητα αλλά ίδια πειράματα τύχης μία φορά.
3. Η πιθανότητα p της πραγματοποίησης του ενδεχομένου A παραμένει σταθερή στις n επαναλήψεις του πειράματος τύχης ή είναι ίδια και στα n πειράματα που θα εκτελεστούν μία φορά.

Άσκηση 202η

Μια βιομηχανία παραγωγής Usb sticks στην Κίνα υποστηρίζει ότι μόνο το **10%** των προϊόντων της μαζικής παραγωγής είναι εκτός των προδιαγραφών της (ελαττωματικό). Παίρνουμε λοιπόν για έλεγχο **10** μονάδες από τα Usb sticks. Αν σε αυτές βρεθούν **2** ή περισσότερες ελαττωματικές μονάδες του προϊόντος, τότε απορρίπτουμε τον ισχυρισμό της βιομηχανίας.

Α) Ποια είναι η πιθανότητα να απορρίψουμε τον ισχυρισμό αδίκως;

Απάντηση:

Πρόκειται για ένα πείραμα δοκιμών Bernoulli με n=10 και p=10%=0,1. Θα απορρίψουμε αδίκως τον ισχυρισμό της βιομηχανίας, αν παρόλο που είναι p=0,1 ωστόσο στο δείγμα βρεθούν 2, 3, 4 ή και μέχρι 10 ελαττωματικές μονάδες.

Με βάση τα παραπάνω μπορούμε να θεωρήσουμε ότι η Διωνυμική Κατανομή με παραμέτρους n=**10** και p=**0,1**είναι το πιθανοτικό πρότυπο για τον αριθμό των ελαττωματικών ανά έλεγχο που θα παρατηρήσουμε. Πιο συγκεκριμένα:

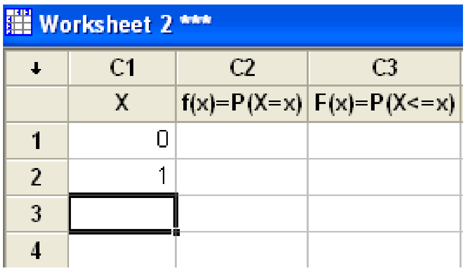
Η τ.μ. X ~ B(10, 0.1) δηλαδή η κατανομή B(10, 0.1) είναι το πιθανοτικό πρότυπο που αποδίδουμε στον πληθυσμό των παρατηρήσεων του αριθμού των ελαττωματικών usbs.

Για να απορρίψουμε τον ισχυρισμό της βιομηχανίας πρέπει να υπολογίσουμε την πιθανότητα $P(X \geq 2)$.

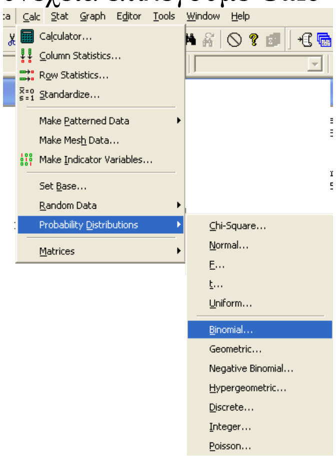
Για τον υπολογισμό των παραπάνω με το MINITAB 15 εργαζόμαστε ως εξής:

1. Ανοίγουμε ένα νέο φύλλο εργασίας (workbook)

2. Στη στήλη C1 καταγράφουμε: X και ακολούθως τις τιμές **0**, και **1**.
3. Στη στήλη C2 καταγράφουμε: $f(x) = P(X = x)$
4. Στη στήλη C3 καταγράφουμε: $F(x) = P(X \leq x)$

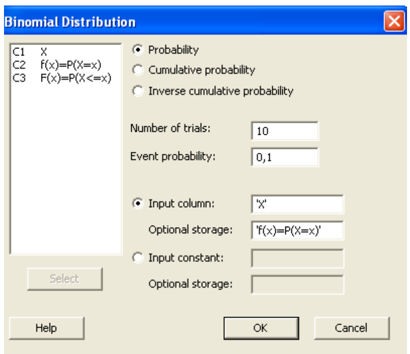


5. στην συνέχεια επιλέγουμε **Calc>Probability Distribution>Binomial**

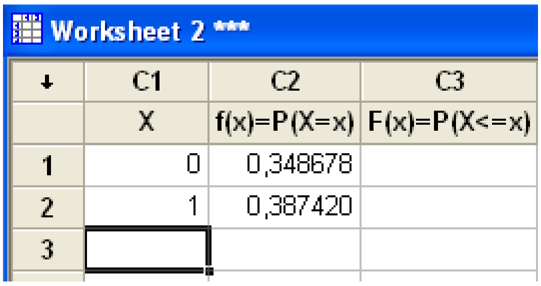


και στην καρτέλα που εμφανίζεται:

6. Κάνουμε κλικ στο **Probability** (για τον υπολογισμό μεμονωμένων πιθανοτήτων).
7. Εισάγουμε στο Number of trial: **10** (συμπληρώνεται η τιμή της παραμέτρου n)
8. Εισάγουμε στο Event Probability: **0,10** (συμπληρώνεται η τιμή της παραμέτρου p)
9. Κάνουμε κλικ στο **Input Column** και θέτουμε Input Column: **X** (ή καταγραφή: **C1) και** Optional Storage: $'f(x)'$ (ή καταγραφή: **C2)** και επιλέγουμε **OK**



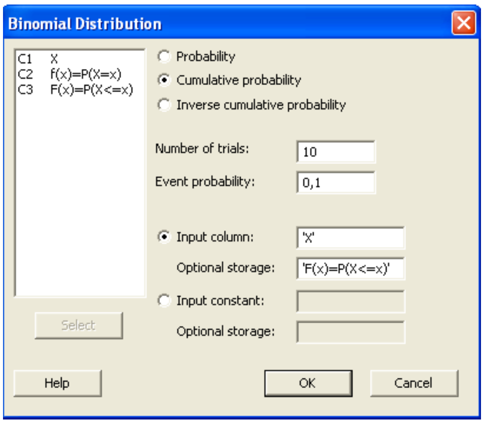
Τα αποτελέσματα είναι:



Worksheet 2 ***

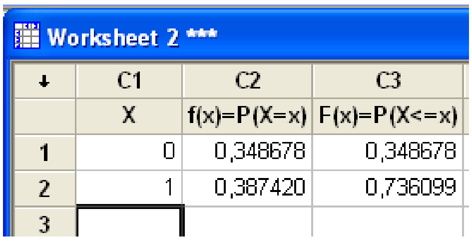
↓	C1	C2	C3
	X	f(x)=P(X=x)	F(x)=P(X<=x)
1	0	0,348678	
2	1	0,387420	
3			

10. Στην συνέχεια επιλέγουμε **Edit→Edit last dialog** και στην καρτέλα που εμφανίζεται:



11. Επιλέγουμε **Cumulative Probability** (για τον υπολογισμό αθροιστικών πιθανοτήτων) και στο Optional Storage: $'F(x) = P(X \leq x)'$ (ή καταγραφή: **C3)** και επιλέγουμε **OK.**

Με ctrl+B μπορούμε να ρυθμίσουμε την ακρίβεια των δεκαδικών. Τα αποτελέσματα που προκύπτουν είναι:



Worksheet 2 ***

↓	C1	C2	C3
	X	f(x)=P(X=x)	F(x)=P(X<=x)
1	0	0,348678	0,348678
2	1	0,387420	0,736099
3			

Σημειώνεται ότι το Minitab δε μπορεί να δώσει απ' ευθείας τη λύση του συγκεκριμένης άσκησης:
Με βάση τον πίνακα όμως μπορεί να υπολογισθεί όμως εύκολα η

$$\boldsymbol{P(X \geq 2) = 1 - [P(X=0) + P(X=1)] = \ldots = 0.264 \text{ ή } 26.4\%}$$

Ή ακόμα πιο εύκολα:

$$\boldsymbol{P(X \geq 2) = 1 - P(X \leq 1) = 1 - [0{,}736] = 0{,}264 \text{ ή } 26{,}4\%}$$

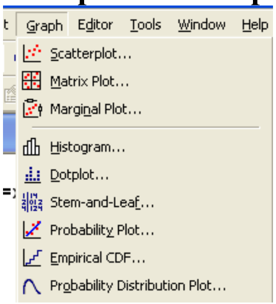
Τέλος μπορούμε με την χρήση του MINITAB 15 να κάνουμε την γραφική παράσταση της σ.π. της παραπάνω διωνυμικής κατανομής **B(10, 0.1).** Αυτό γίνεται ως εξής:

1ος τρόπος

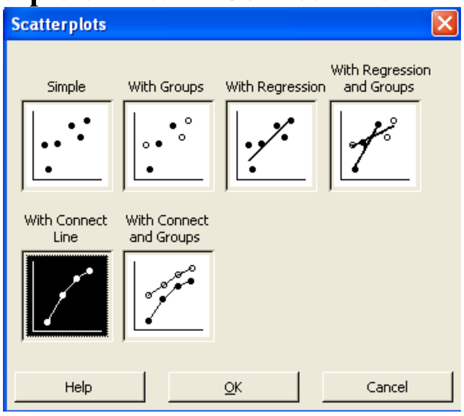
1. Εισάγουμε τις τιμές για το X=1,2 …10 στην C1 στήλη ενώ στην C2 στήλη υπολογίζουμε την $'f(x)'$ με τον τρόπο που είδαμε πριν. Προκύπτουν τα εξής:

↓	C1	C2
	X	f(x)=P(X=x)
1	0	0,348678
2	1	0,387420
3	2	0,193710
4	3	0,057396
5	4	0,011160
6	5	0,001488
7	6	0,000138
8	7	0,000009
9	8	0,000000
10	9	0,000000
11	10	0,000000

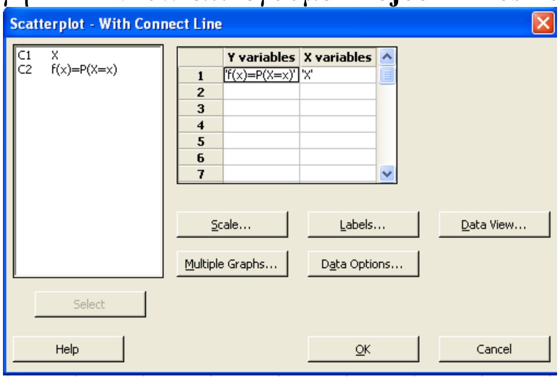
2. Στην συνέχεια επιλέγουμε **Graph→ Scatterplot**



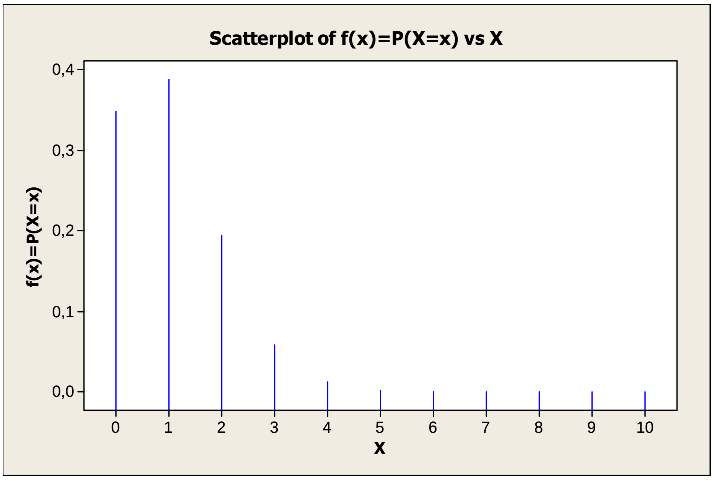
3. Επιλέγουμε **Scatterplots → With Connect line**



4. Θέτουμε τις τιμές X στον άξονα –χ και τις τιμές της σ.π. στον άξονα –Y ενώ από την επιλογή **Data View** επιλέγουμε **Project Lines** και μετά **OK**.

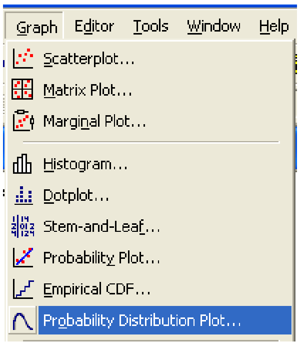


Η γραφική που προκύπτει κάνοντας κλικ στο παράθυρο του Scatterplot είναι

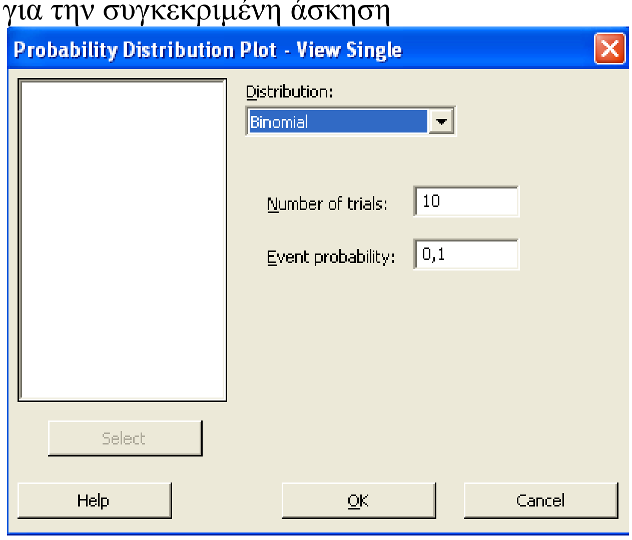


$2^{ος}$ τρόπος

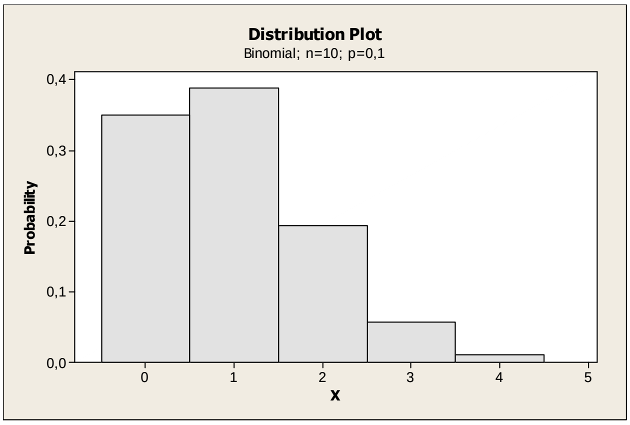
Το MINITAB 15 μας δίνει την δυνατότητα να δημιουργήσουμε οποιαδήποτε γραφική παράσταση σ.π. κατανομής θέλουμε από την επιλογή **Graph→Probability Distribution Plot**



Στην συνέχεια επιλέγοντας από την πτυσσόμενη λίστα του Distribution την Binomial κατανομή και θέτοντας στο **Number of trials = 10** και στο **Event probability = 0.1** για την συγκεκριμένη άσκηση



και επιλέγοντας **OK** έχουμε τελικά:



Από τις γραφικές παραστάσεις της σ.π. της διωνυμικής κατανομής μας προκύπτει ότι η κατανομή είναι λοξή προς τα αριστερά.

Εργασία:

Στην παραπάνω άσκηση ποια είναι η πιθανότητα να απορρίψουμε τον ισχυρισμό, αν και στην πραγματικότητα το 20% της παραγωγής των προϊόντων της παραγωγής είναι ελαττωματικά;