MINITAB tutorial TA session

Yuanyuan Gao

Table of contents

1. Calculator, basic statistics, histogram, boxplot

- Problem #2 in Exam #1:
- Open the MINITAB data set PulseA.MTW

Open the MINITAB data set PulseA.MTW

(a) Obtain side-by-side box plots of Pulse 2(Column C2) versus Activity (Col. C8).
Find:

- Which Activity level (1, 2 or 3) has highest IQR?

  Activity level ; IQR =

- Which Activity level has highest Median?

  Activity level ; Median=

- Are there any outliers? If so, identify:

  Activity level(s) ; high or low side?

  Outlier=

(b) Construct a new column of the difference:

DIFF = Pulse 2 – Pulse 1

Obtain the following descriptive statistics of DIFF classified by sex (col. C5):

Variable	Sex	N	Mean	St Dev	IQR	Coeff Var
DIFF	Female					
	Male					

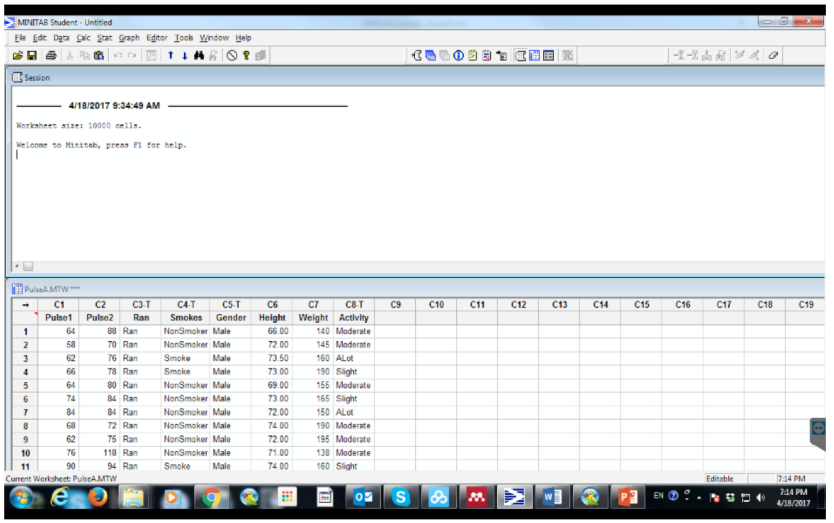
(c) For the over-all weight (Col. C7), find approx. values of

51st percentile =

63rd percentile =

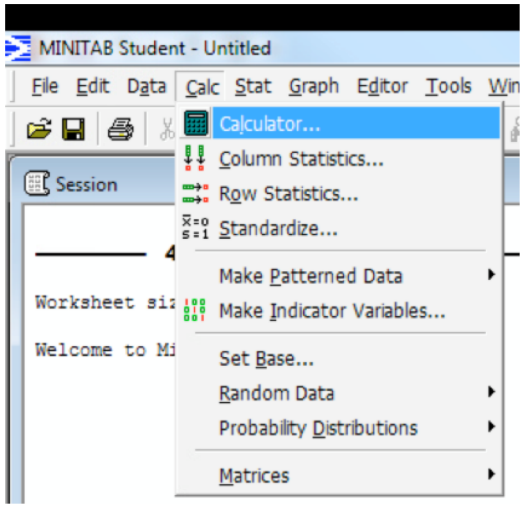
(d) Find Prob[DIFF > 12]=

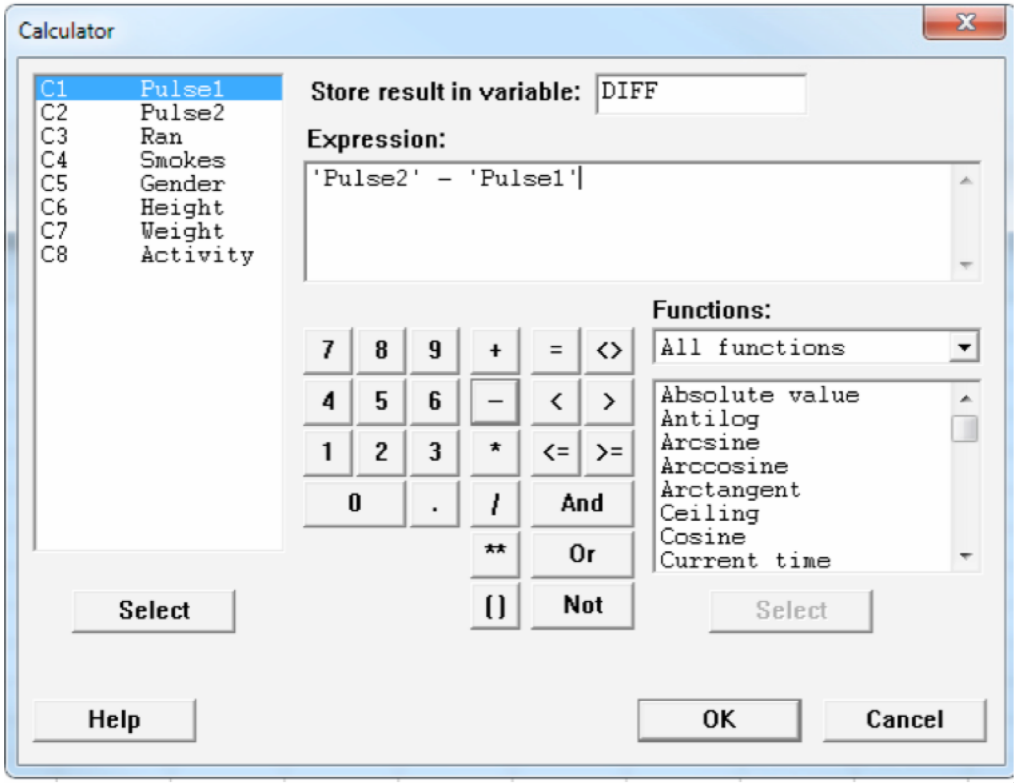
(e) Check the normality of Weight (Col C7): p-value = _____.



Calculator

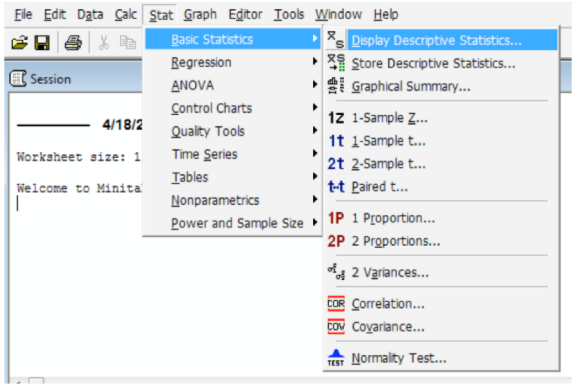
Calc->Calculator



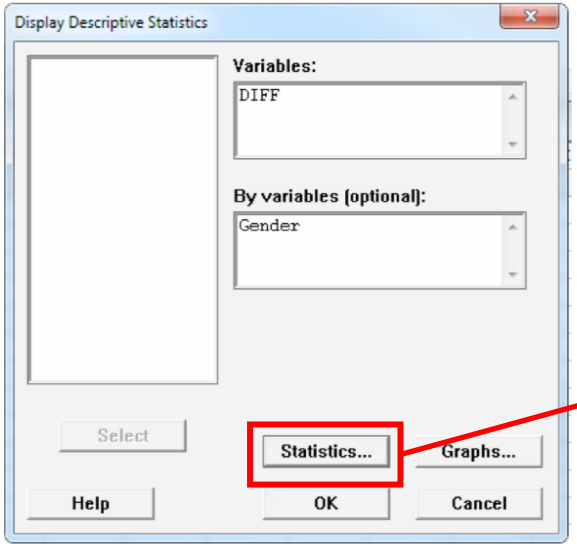


Basic statistics

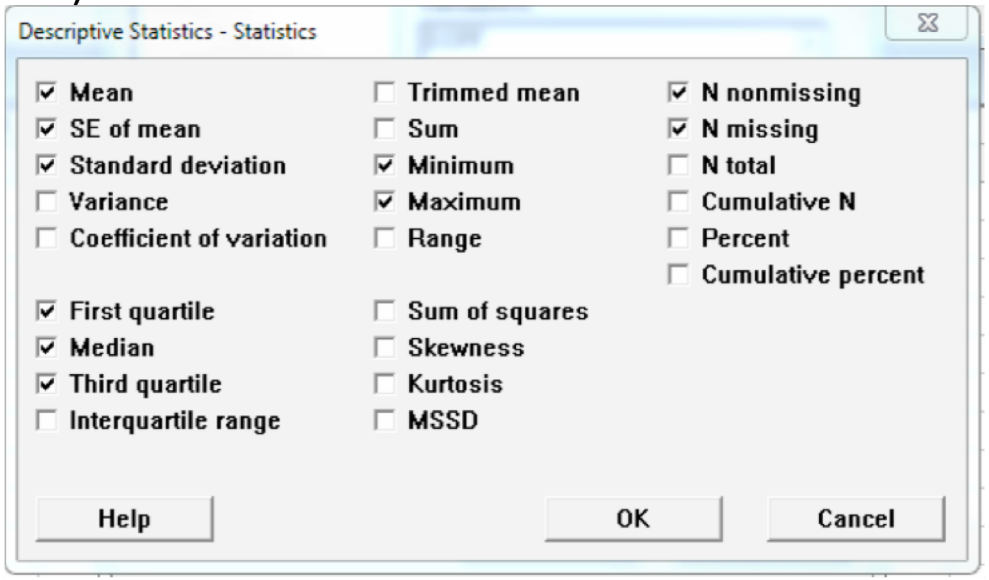
1. Stat->Basic Statistics->Display Descriptive Statistics



2. Input variables and By variables(optional)



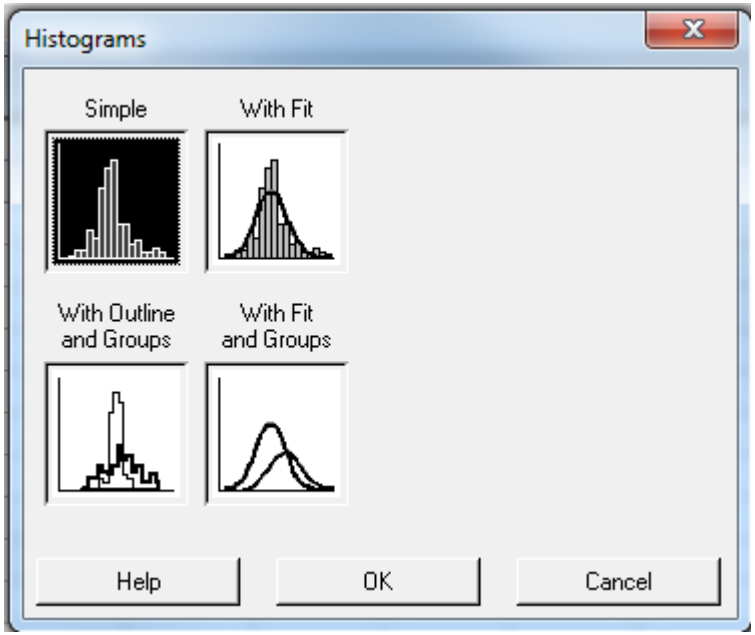
3. In Statistics, we can choose information we want

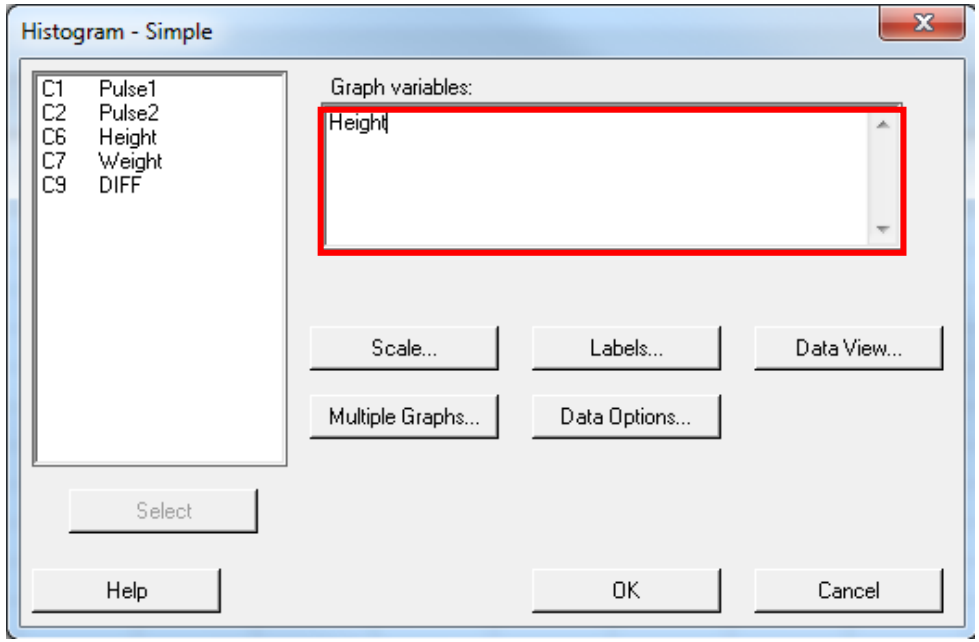


Histogram

- Graph-> Histogram

Input the variable

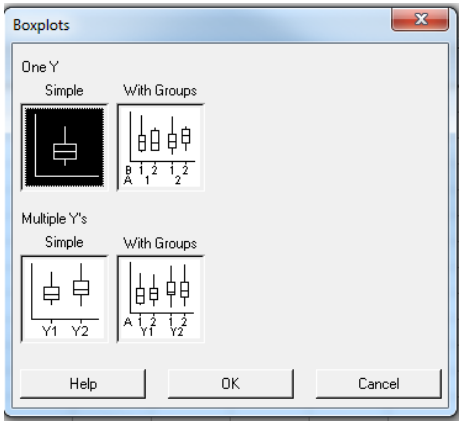


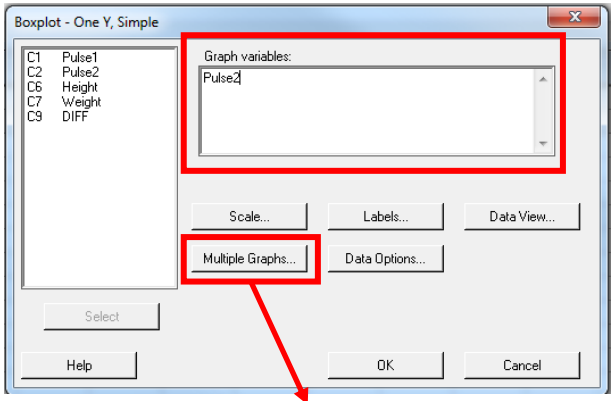


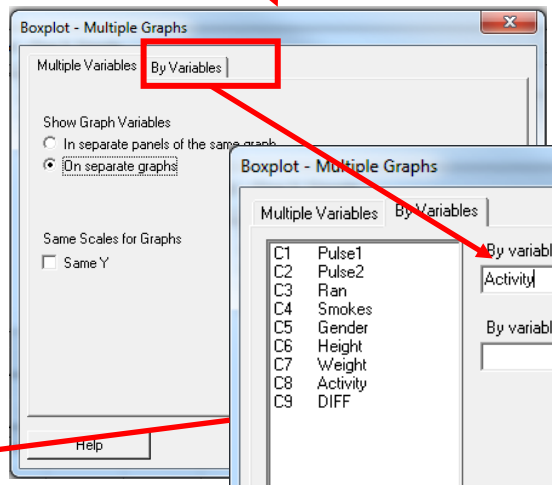
Obtain side-by-side boxplot of Pulse2 vs. Activity

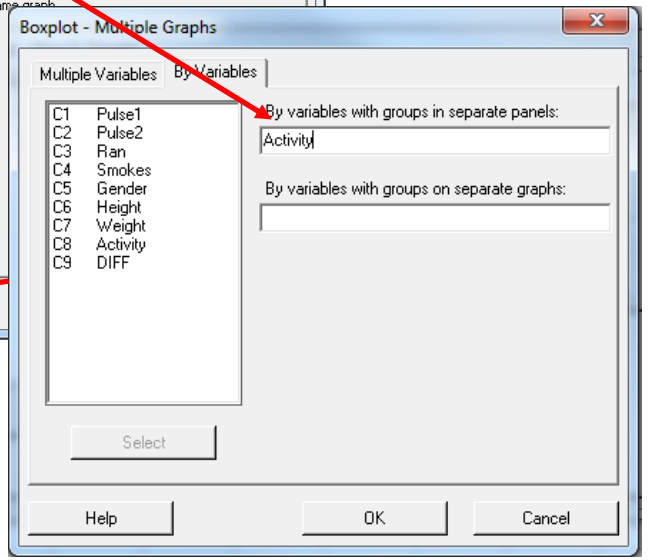
Boxplot

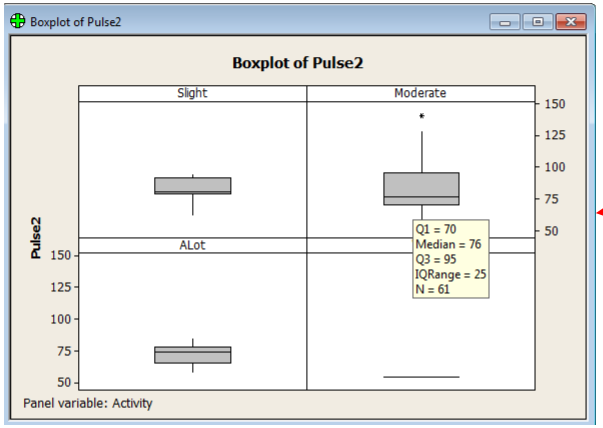
- Graph-> Boxplots







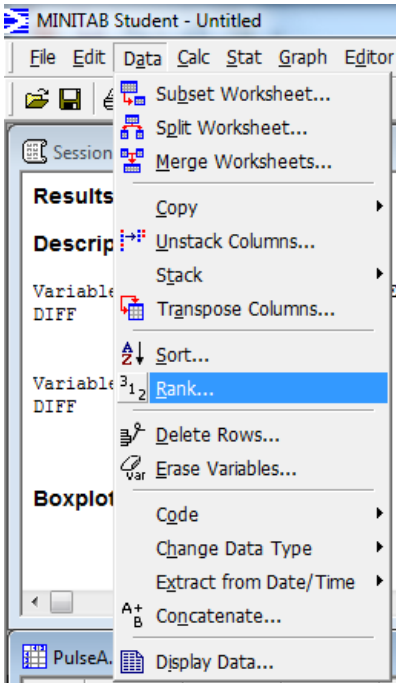




When the mouse is over the plot, Q1 etc. will show up.
The same to outliers.

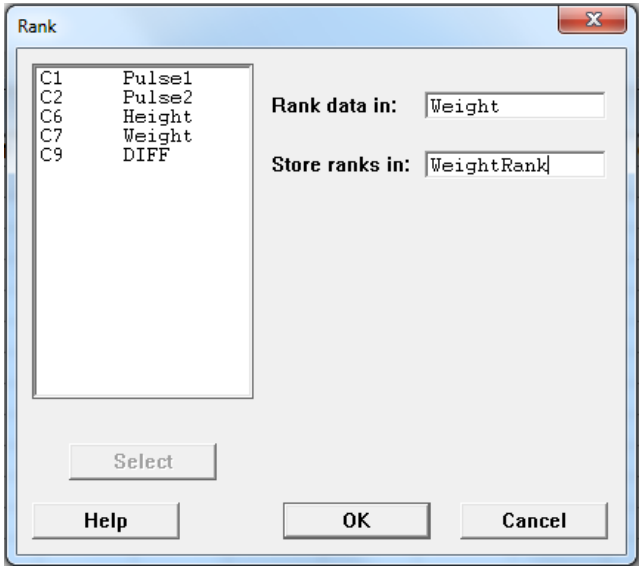
Percentile in a column of data

Data-> Rank



51st Percentile in 'Weight'= ?

Total # in 'Weight' is 92. The 51st percentile is the 92*51%=46.92th number. Find a number in WeightRank close to 46.92, the answer is the according Weight.



→	C1	C2	C3-T	C4-T	C5-T	C6	C7	C8-T	C9	C10
	Pulse1	Pulse2	Ran	Smokes	Gender	Height	Weight	Activity	DIFF	WeightRank
1	64	88	Ran	NonSmoker	Male	66.00	140	Moderate	24	39.5
2	58	70	Ran	NonSmoker	Male	72.00	145	Moderate	12	45.0

2. Probability distribution

- Example: Problem 3 in Exam #1

Problem 3

[A] Using MINITAB, evaluate the following:

a. For X~Bin(n=15,p=.45)

p[x = 7] = _____.

p[$x \leq 6$] = _____.

p[x>5] = _____.

b. For X~Normal with $\mu = 11, \sigma = 2.5$

P[x < 8] = _____.

81st percentile =_____ .

c. For Y ~ Poisson, with $\lambda = 10$

P[x > 8]= _____ .

P[x = 5] =_____.

d. For X~exponential, with $\lambda = 5$

Mean = _____.

Median = _____.

Probability distribution

- Discrete
  - Binomial
  - Poisson
- Continuous
  - Normal
  - Exponential
  - Gamma

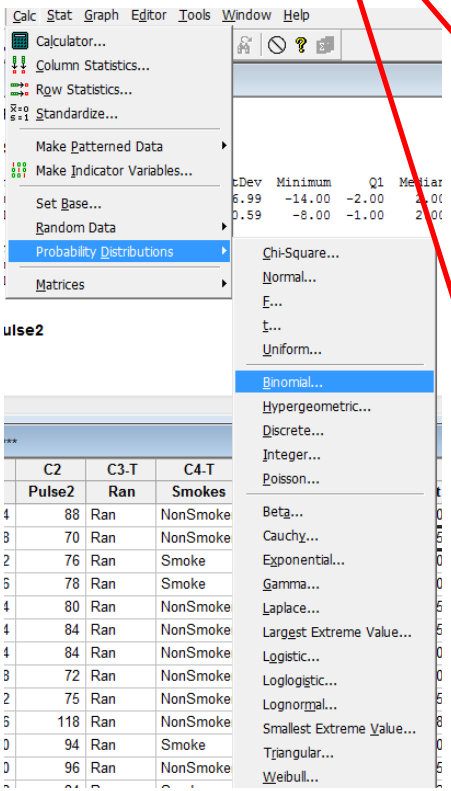
Binomial

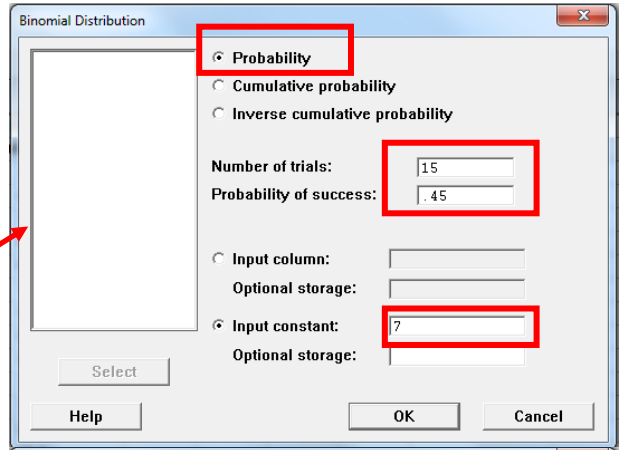
a. For X~Bin(n=15,p=.45)

p[x = 7] = _____.

p[$x \leq 6$] = _____.

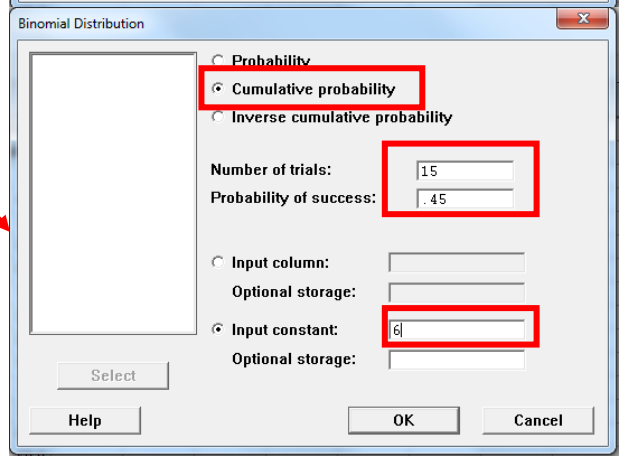
p[x>5] = 1-p[x<=5]





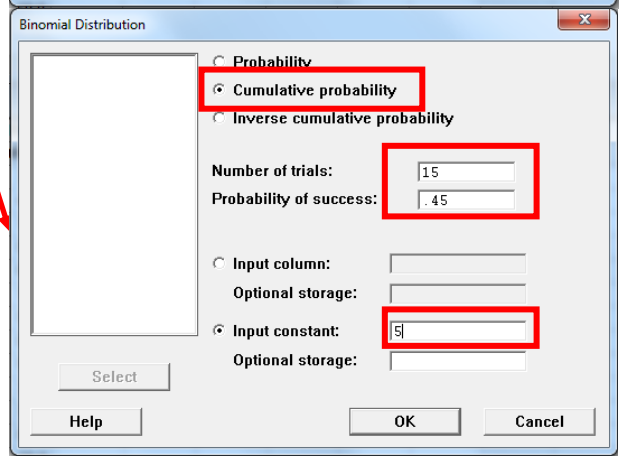
Binomial with n = 15 and p = 0.45

x	P(X = x)
7	0.201344



Binomial with n = 15 and p = 0.45

x	P(X <= x)
6	0.452160



Binomial with n = 15 and p = 0.45

x	P(X <= x)
5	0.260760

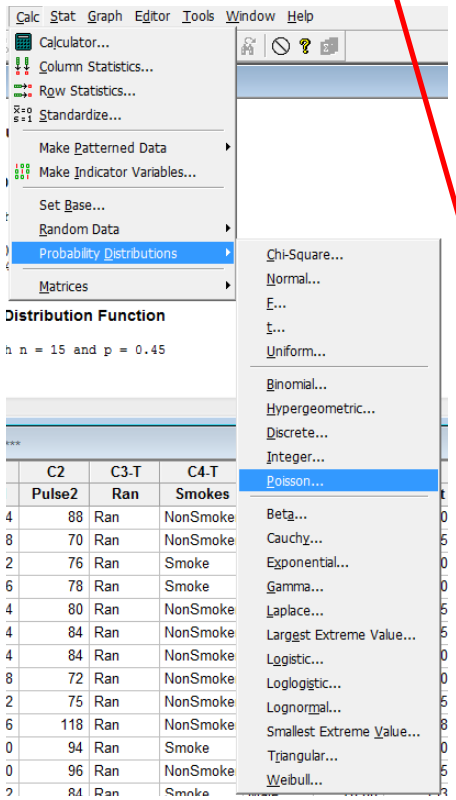
1 - 0.260760 = 0.73924

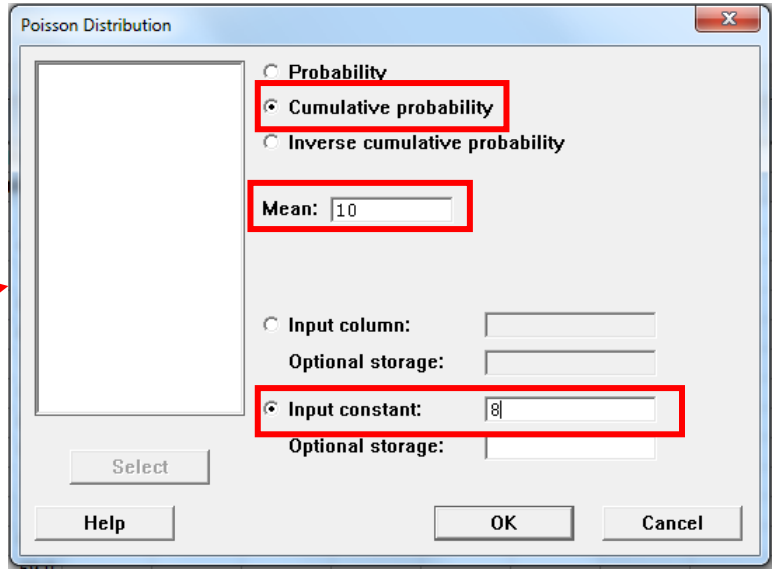
Poisson

c. For Y ~ Poisson, with $\lambda = 10$

P[x > 8]= 1-p[x<=8] .

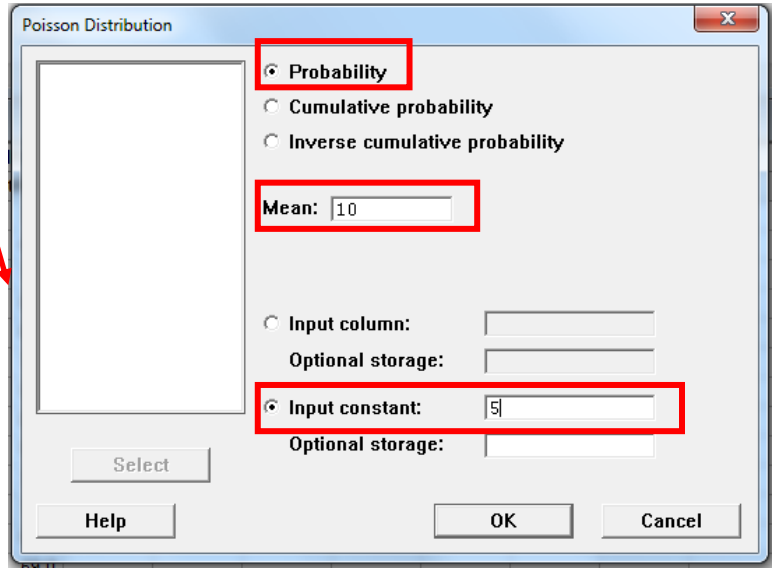
P[x = 5] = _____.






```
Poisson with mean = 10

x   P( X <= x )
8     0.332820
```

1 – 0.332820 = 0.66717



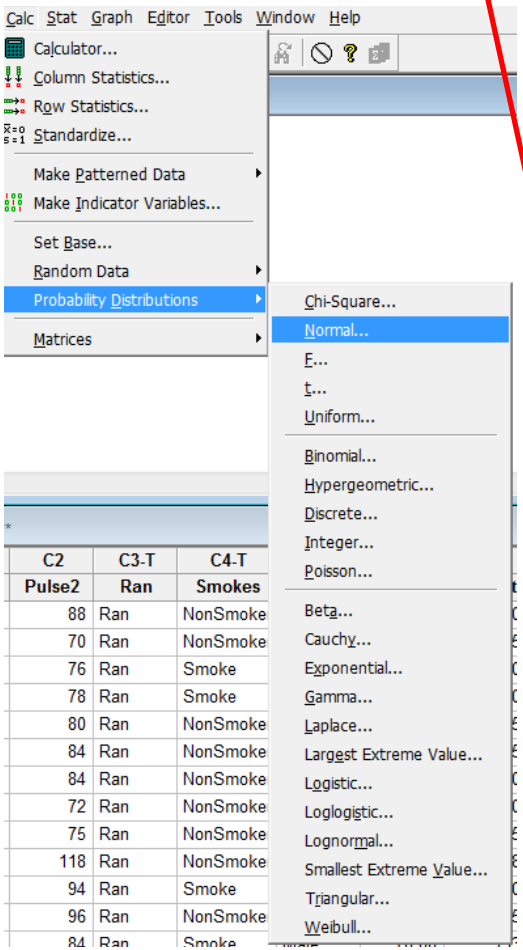

```
Poisson with mean = 10

x   P( X = x )
5    0.0378333
```

Normal

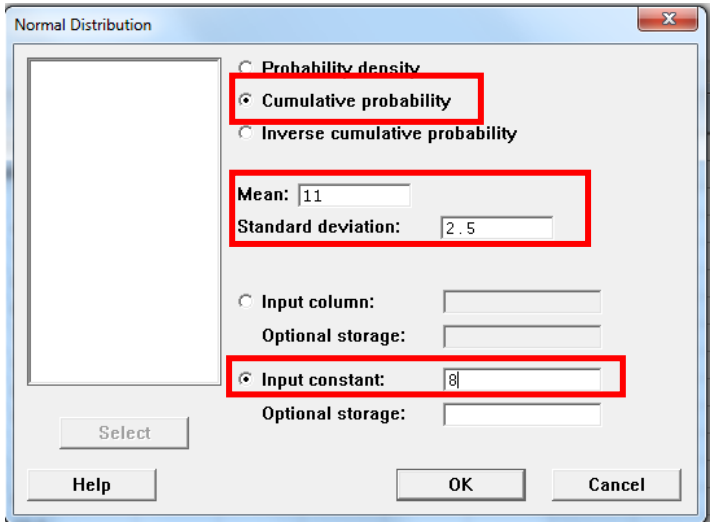
b. For X~Normal with $\mu = 11, \sigma = 2.5$

P[x < 8] = _____.

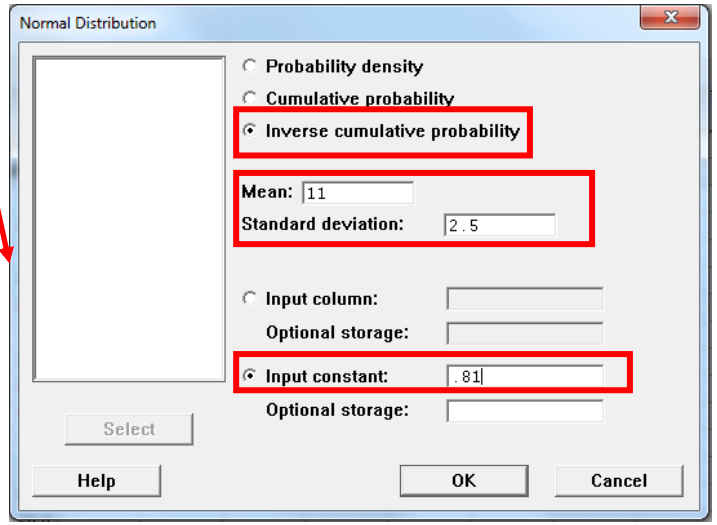
81st percentile =_____ .



C2	C3-T	C4-T
Pulse2	Ran	Smokes
88	Ran	NonSmoke
70	Ran	NonSmoke
76	Ran	Smoke
78	Ran	Smoke
80	Ran	NonSmoke
84	Ran	NonSmoke
84	Ran	NonSmoke
72	Ran	NonSmoke
75	Ran	NonSmoke
118	Ran	NonSmoke
94	Ran	Smoke
96	Ran	NonSmoke
84	Ran	Smoke




```
Normal with mean = 11 and standard deviation = 2.5

x   P( X <= x )
8       0.115070
```



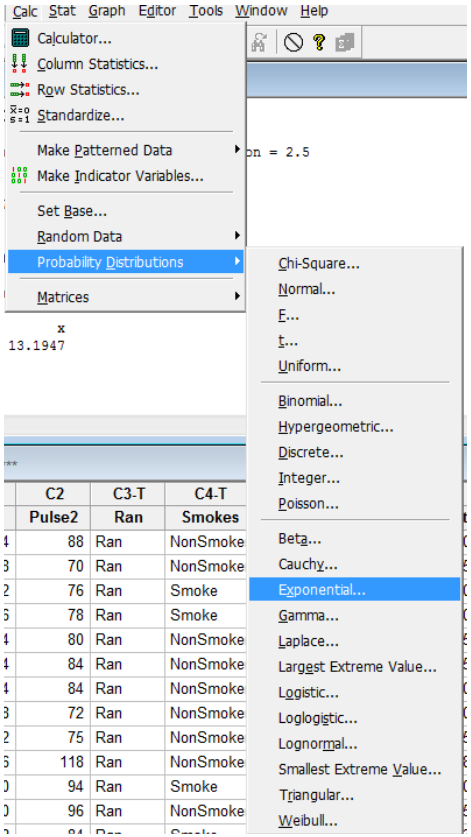

```
Normal with mean = 11 and standard deviation = 2.5

P( X <= x )          x
       0.81   13.1947
```

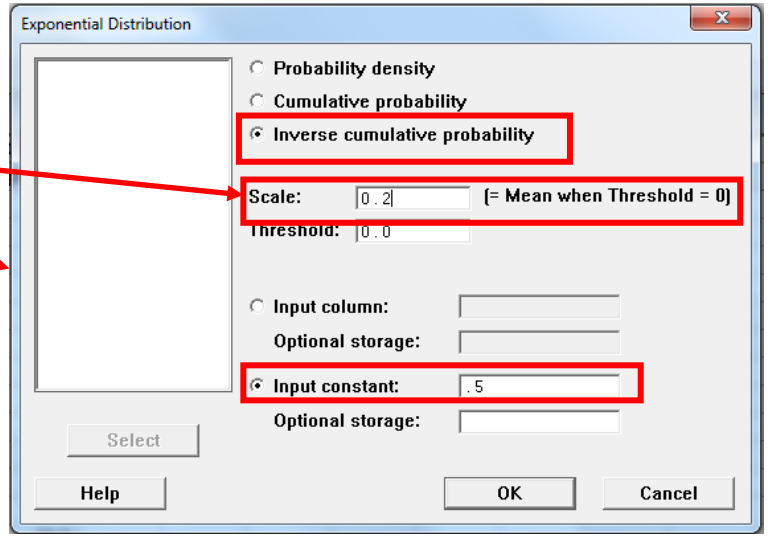
Exponential

d. For X~exponential, with $\lambda = 5$

Mean = 1/5 = 0.2.

Median = _____.






```
Exponential with mean = 0.2

P( X <= x )          x
        0.5   0.138629
```

3. Random data, column statistics, row statistics, monte carlo

- For example Prob. 3 in Exam 2

[A] The lifetime of three lamps is exponentially distributed with means 120, 160 and 130 hours respectively. If they are placed in parallel in a system, estimate (using Monte Carlo simulation with 500 iterations), the

(a) Prob[system lifetime > 150] = _____.

(b) Is the probability distribution of system lifetime

Normal?_____.

Exponential? _____ .

Gamma? _____ .

How do you check this? Give p values.

[B] The acceleration g due to gravity is measured by dropping an object and measuring the time t it takes to travel a distance S. Assuming s and t are normally distributed with means and uncertainties as indicated:

$$s = 2.2 \pm 0.01m$$

$$t = 0.67 \pm 0.015s$$

(i) Estimate (using Monte Carlo simulation with 500 iterations), g and its uncertainty.

g = _____ $\pm$ _____ [Hint: s = $\frac{1}{2}gt^2$]

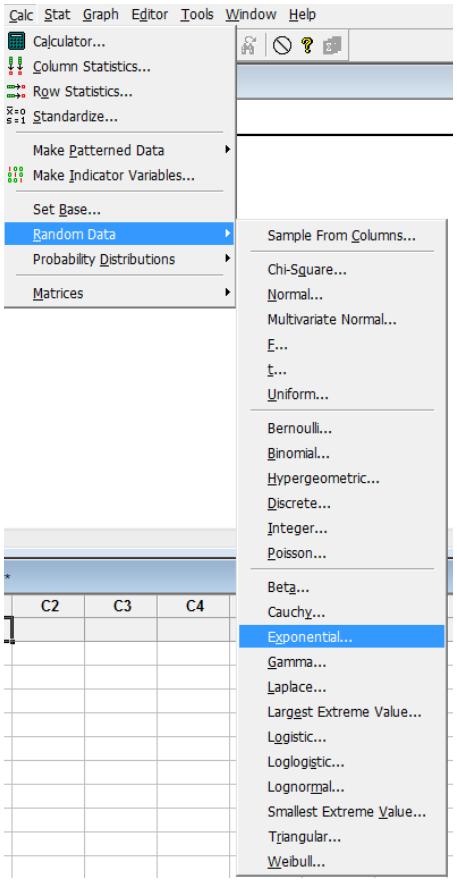
(ii) Run an Exec file (100 iterations) on the mean value of g. Find the average.

Copy the exec file.

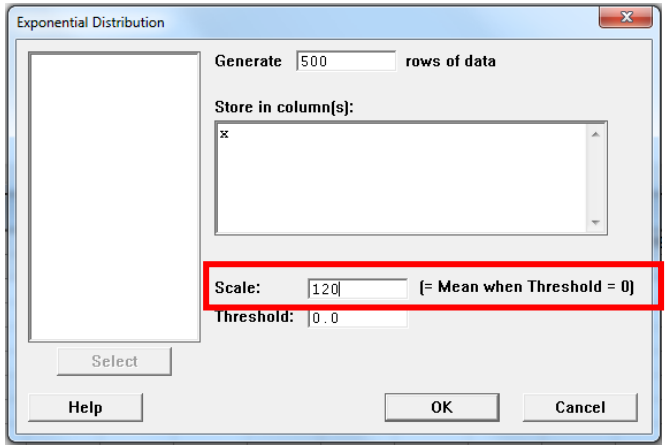
Random data

[A] The lifetime of three lamps is exponentially distributed with means 120, 160 and 130 hours respectively. If they are placed in parallel in a system, estimate (using Monte Carlo simulation with 500 iterations), the

(a) Prob[system lifetime > 150] = _____.



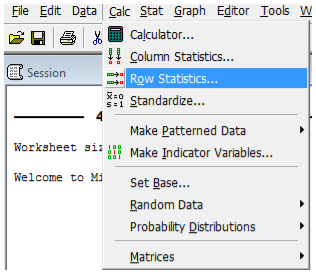
Generate x,y,z columns of data

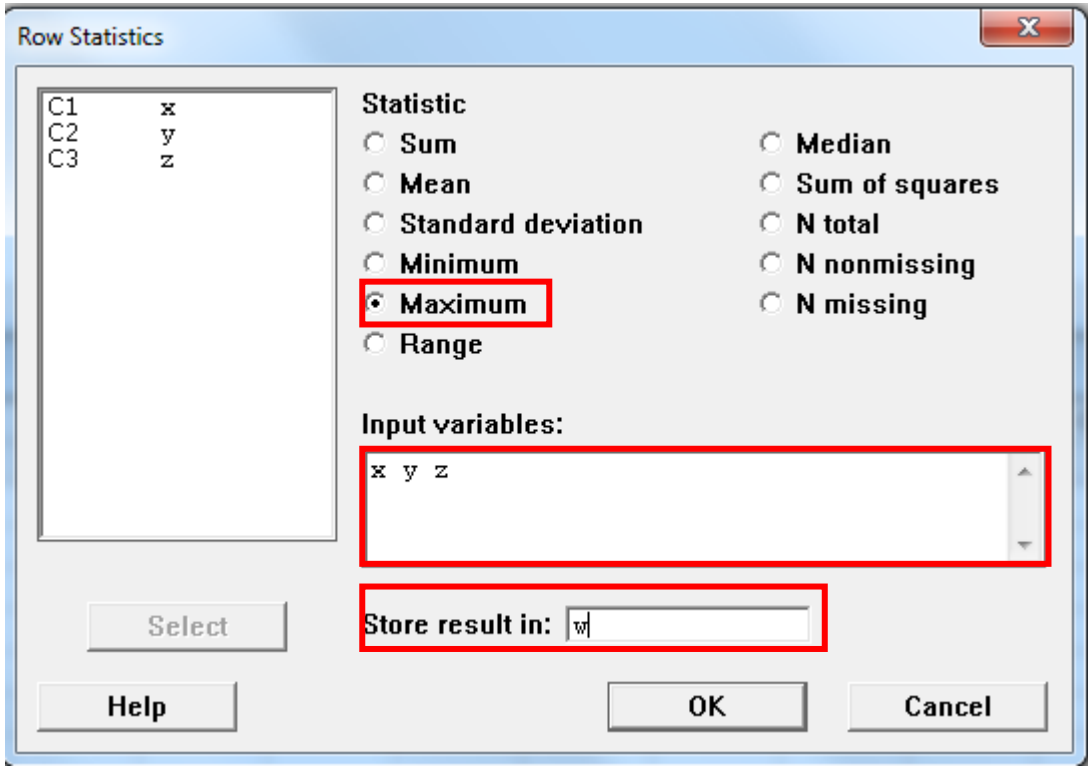


C1	C2	C3
x	y	z
198.169	68.984	94.73
188.082	1.885	77.73
15.990	279.901	132.09
265.590	21.292	135.30
227.518	126.934	86.40
256.802	285.487	27.64
138.813	17.422	163.50
103.351	62.350	14.05
236.884	75.233	37.44
112.131	19.700	161.52

Row statistics

Calc-> Row Statistics



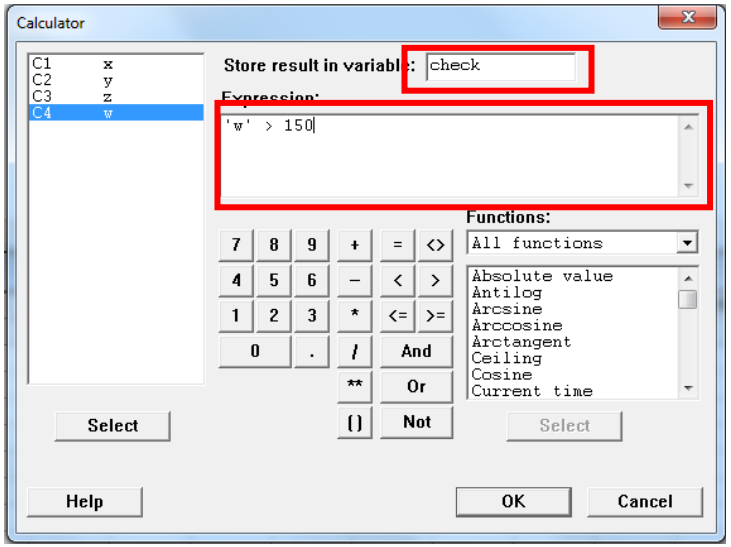


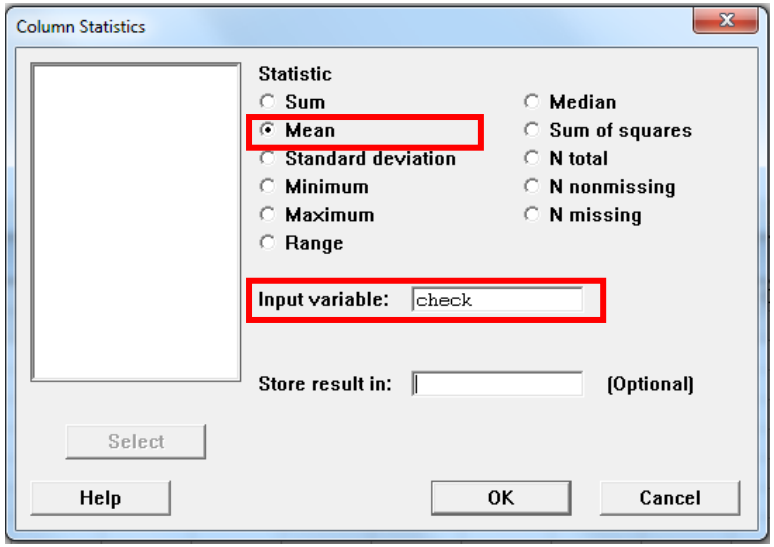
x	y	z	w
198.169	68.984	94.73	198.17
188.082	1.885	77.73	188.08
15.990	279.901	132.09	279.90
265.590	21.292	135.30	265.59
227.518	126.934	86.40	227.52
256.802	285.487	27.64	285.49
138.813	17.422	163.50	163.50
103.351	62.350	14.05	103.35

Column statistics

[A] The lifetime of three lamps is exponentially distributed with means 120, 160 and 130 hours respectively. If they are placed in parallel in a system, estimate (using Monte Carlo simulation with 500 iterations), the

(a) Prob[system lifetime > 150] = _____.





Mean of check

Mean of check = 0.7

Monte Carlo

[B] The acceleration g due to gravity is measured by dropping an object and measuring the time t it takes to travel a distance S. Assuming s and t are normally distributed with means and uncertainties as indicated:

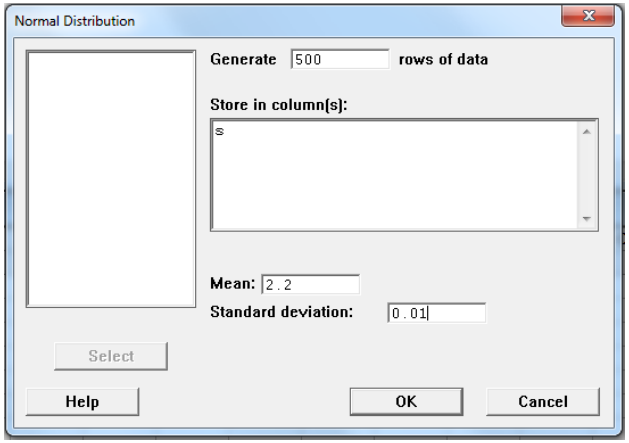
$$s = 2.2 \pm 0.01m$$

$$t = 0.67 \pm 0.015s$$

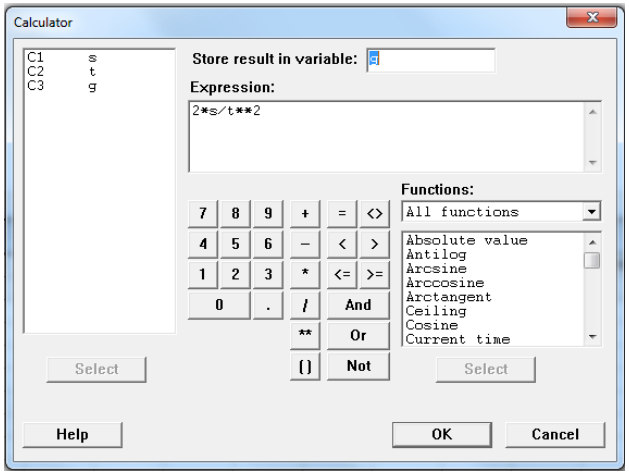
(i) Estimate (using Monte Carlo simulation with 500 iterations), g and its uncertainty.

g = _____ ± _____ [Hint: $s = \frac{1}{2}gt^2$]

1. Generate 2 columns for ‘s’ and ‘t’

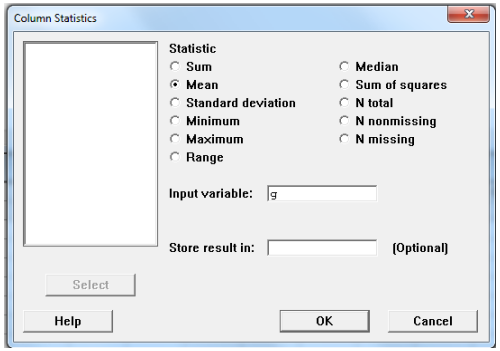


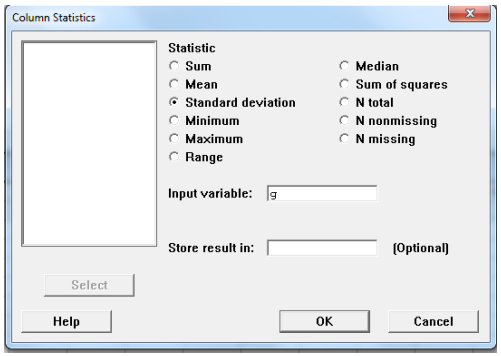
2. Calculate ‘g’



s	t	g
2.19751	0.666008	9.9084
2.19201	0.665263	9.9057
2.20787	0.691967	9.2222
2.21312	0.670899	9.8338
2.21404	0.677959	9.6340
2.20907	0.692054	9.2248
2.19018	0.669479	9.7732
2.20587	0.670413	9.8158

3. Column statistic for ‘g’





Mean of g

```
Mean of g = 9.86910
```

Standard Deviation of g

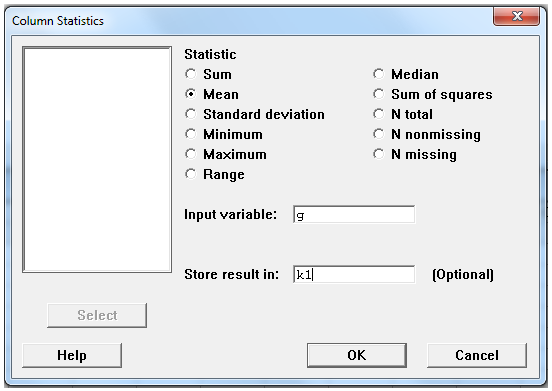
```
Standard deviation of g = 0.469206
```

Monte Carlo

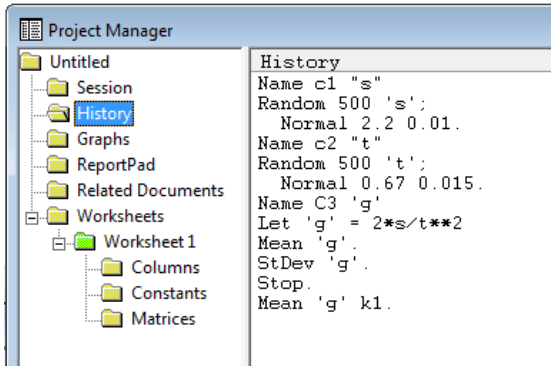
(ii) Run an Exec file (100 iterations) on the mean value of g. Find the average.

Copy the exec file.

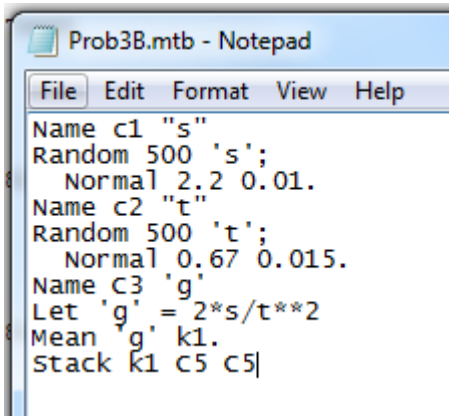
1. Store mean of ‘g’ in k1



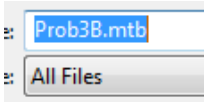
2. Project manager -> History-> Copy all the scribe



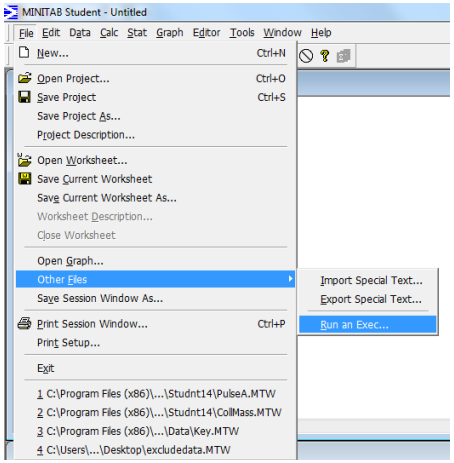
3. Tools-> Notepad-> paste sctibe-> add ’stack k1 C5 C5’



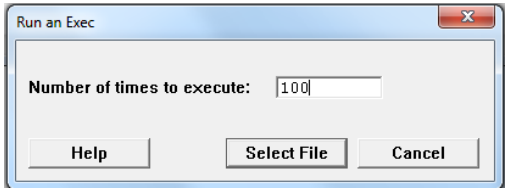
4. Save as ‘Prob3B.mtb’. Choose ‘All Files’



5. File->Other Files->Run an Exec

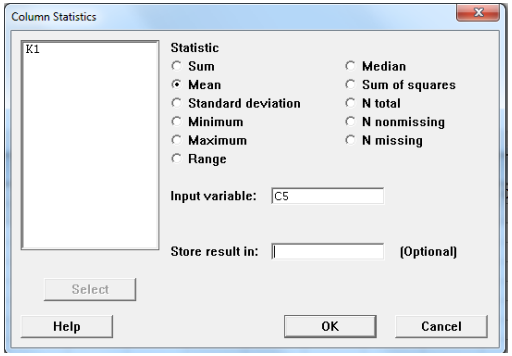


6. Indicate trails number and select ‘Prob3B.mtb’



Monte Carlo

- After running the exec file, calculate the mean of C5



Mean of C5

```
Mean of C5 = 9.81614
```

Another way to generate data - bootstrap

- Prob 6 in Exam 2

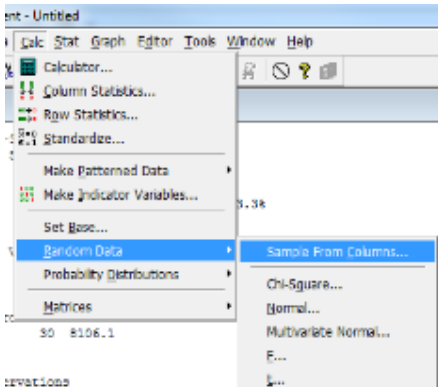
[A] Following sample is taken from an unknown population:

$\{x_i\}$: 10, 19, 22, 13, 9, 16, 12, 24, 17, 23, 18, 14

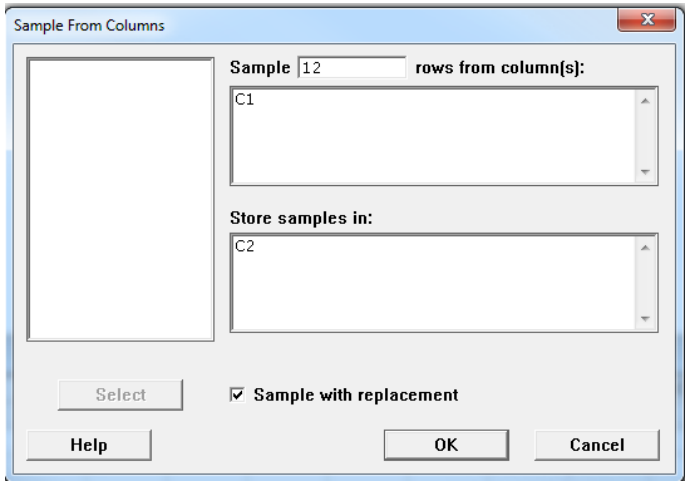
Using an appropriate approach, obtain an estimate of the standard error of the median, $\sigma_{\tilde{x}}$.

C1	C2
10	9
19	9
22	18
13	9
9	23
16	24
12	17
24	9
17	12
23	23
18	19
14	18

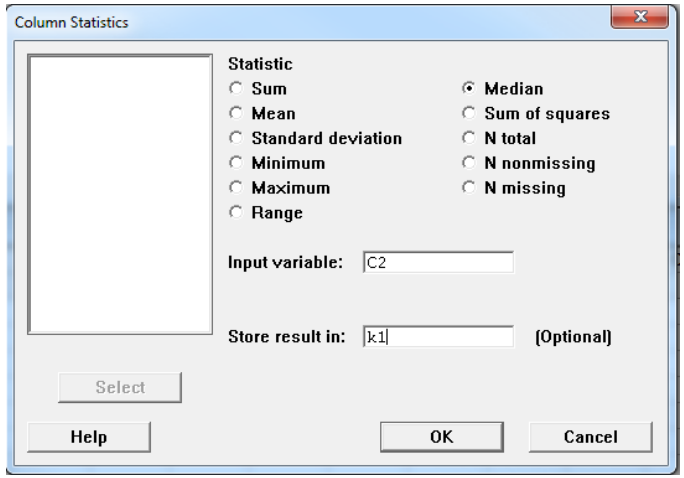
1. Calc-> Random data-> Sample from column



2. Input sample number, input column and output column

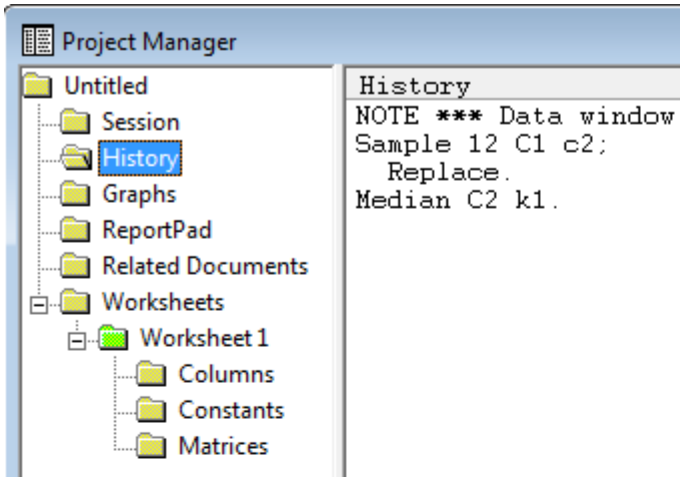


3. Calc -> column statistics-> Median-> Input variable C2 store k1

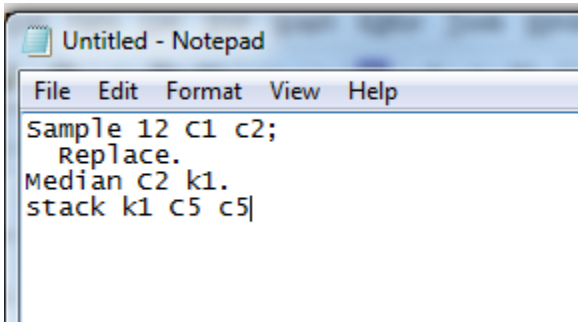


Bootstrap

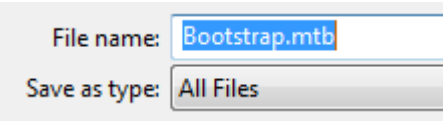
4. Project Manager-> History-> Copy



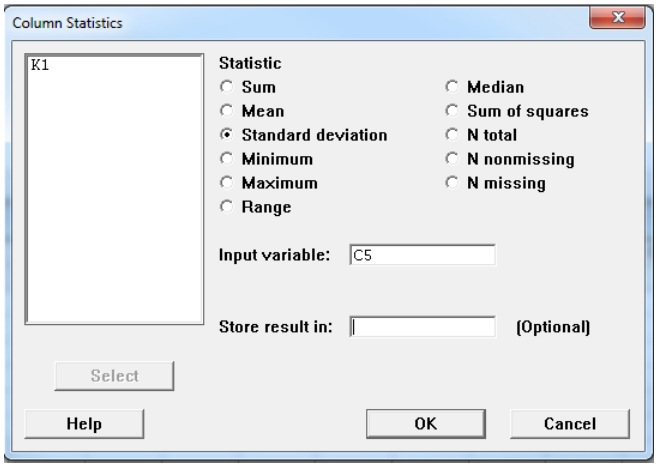
5. Paste it in Notepad and add ’stack k1 C5 C5’



6. Save as ‘Bootstap.mtb’



7. Run the exec file 100 times, then Column statistics-> Standard deviation -> Input C5



Standard Deviation of C5

Standard deviation of C5 = 1.97215

4. t-test, z-test, power analysis, probability plot

- Problem 2 in Exam 2

We want to test if there is a difference in true average bacteria count (number of colonies/ft^3) between carpeted (X_i) and uncarpeted (Y_i) rooms. The following samples were taken in a hospital: [Assume Equal Variance 5]

X_i	15.3	16.4	16.2	15.5	15.0	15.3	14.4	15.1	15.6	15.2	16.3
Y_i	13.4	14.8	13.0	14.1	13.7	15.4	16.3	13.2	14.3	15.3	13.9

[Note: $\sum x_i = 170.3; \sum y_i = 157.4$]

(a) State appropriate hypothesis to be tested:

(b) What type of test would you use?

(c) Give values of

Test statistic =

P-value =

s.d. =

(d) Is H_0 rejected or not @ = .05 ?

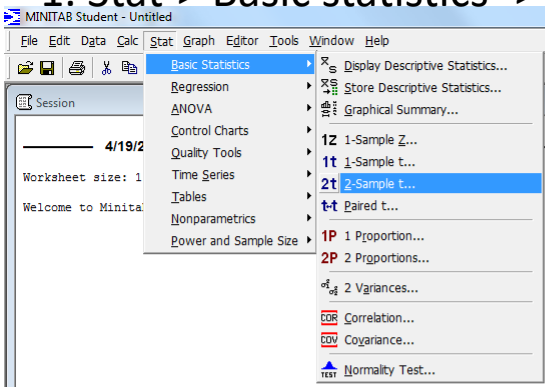
(e) Give a 95% CI on $\mu_x - \mu_y$:

(f) If actually $\mu_x - \mu_y = 1$, how large a sample size would be necessary to test the difference, with a power of .95? (Take $\alpha = .05$)

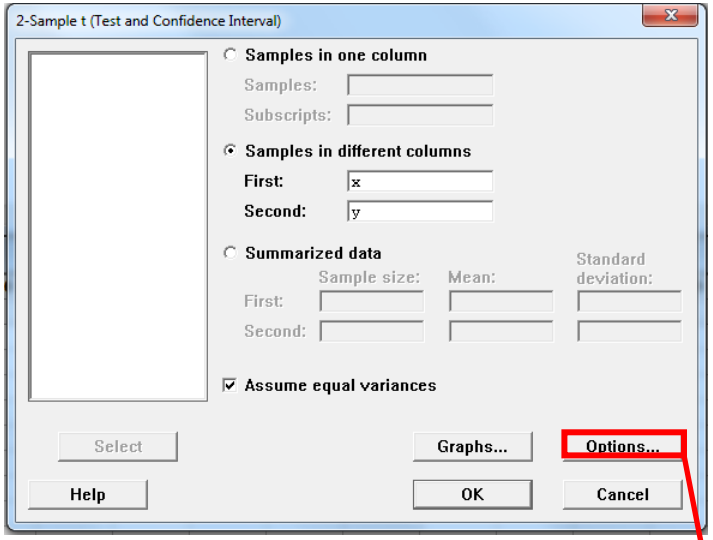
(g) What main assumption underlies your test? Is the assumption 'justified'?

T-test, z-test

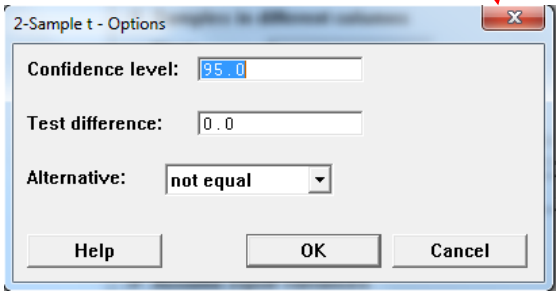
1. Stat-> Basic statistics -> 2-sample t



2. Input samples in different columns



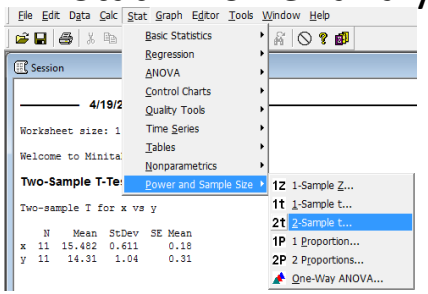
3. Options-> Confidence level, test difference, Alternative



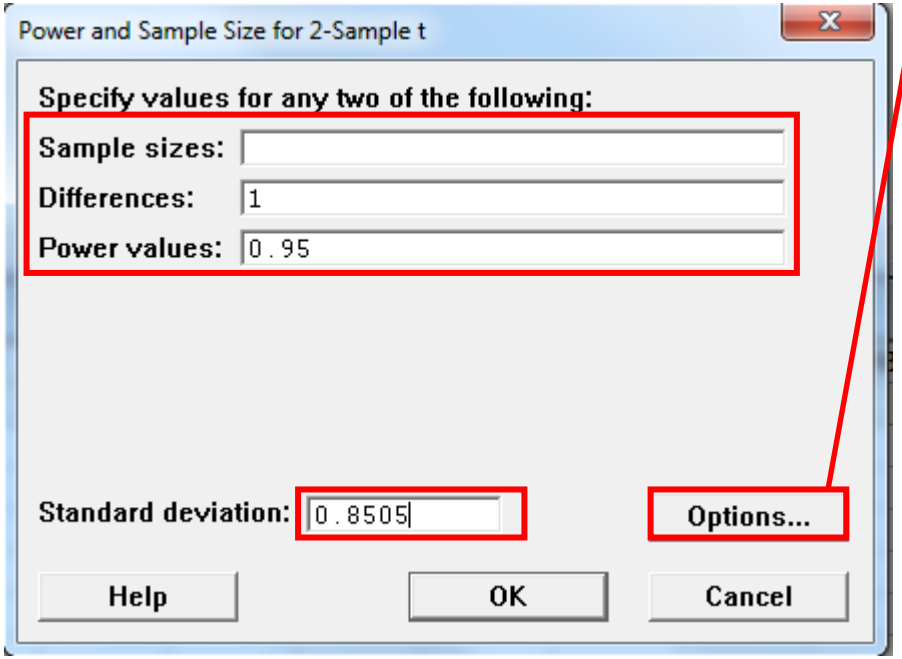

```
Two-sample T for x vs y

    N    Mean  StDev  SE Mean
x  11  15.482  0.611     0.18
y  11   14.31   1.04     0.31


Difference = mu (x) - mu (y)
Estimate for difference:  1.17273
95% CI for difference:  (0.41628, 1.92917)
T-Test of difference = 0 (vs not =): T-Value = 3.23  P-Value = 0.004  DF = 20
Both use Pooled StDev = 0.8505
```

Power analysis

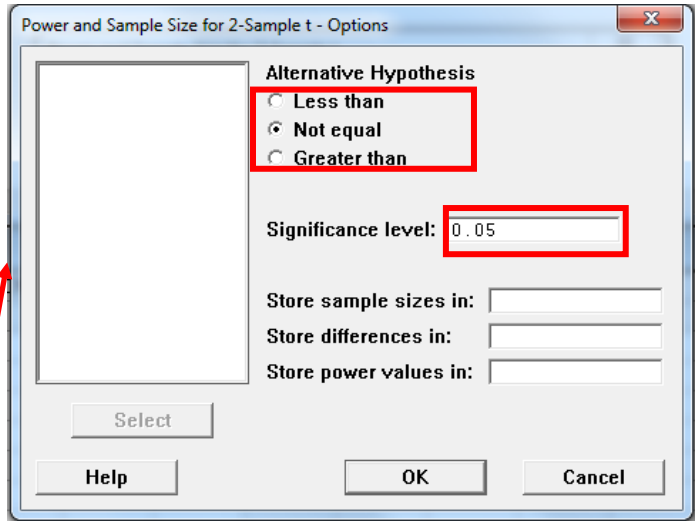
1. Stat-> Power analysis -> 2 sample t test



2. Input differences and power values, and s.d.



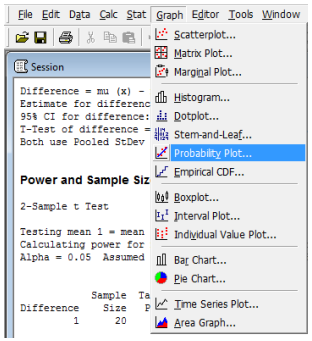
3. Alternative Hypothesis and sig. level

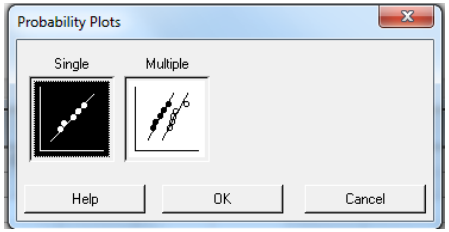



```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 0.8505


            Sample  Target
Difference    Size   Power  Actual Power
         1      20    0.95      0.951810

The sample size is for each group.
```

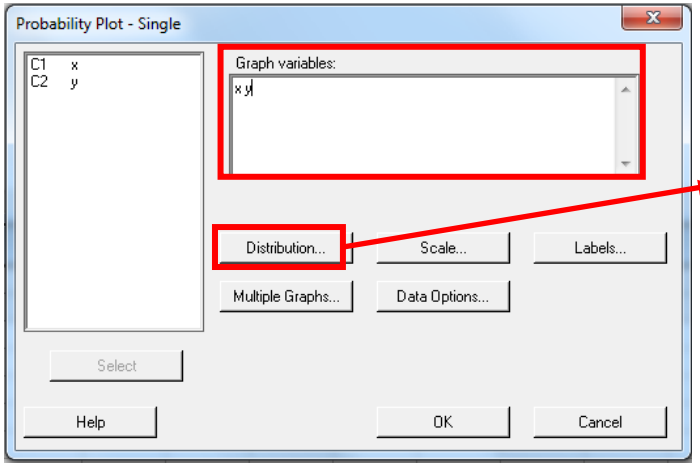
Probability plot

Graph-> Probability Plot

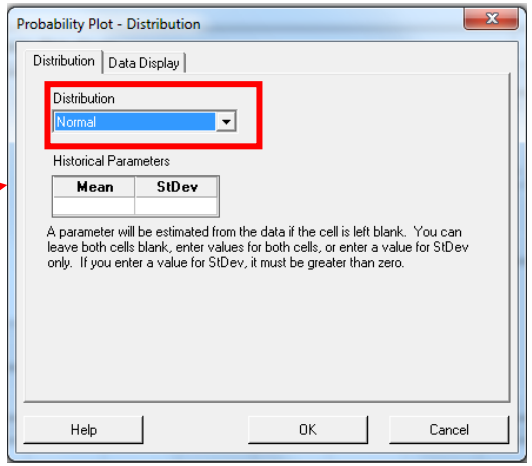


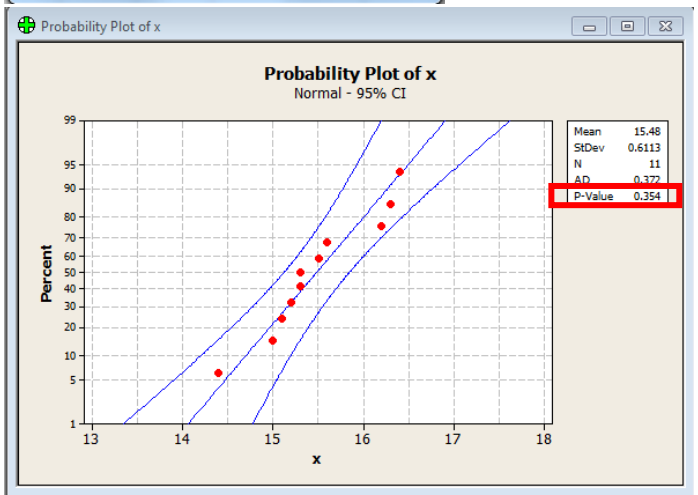


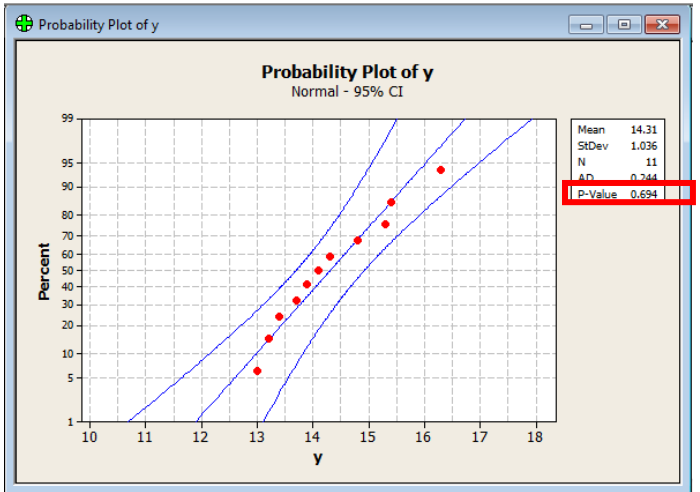
Insert variables



Choose the probability distribution we want to test







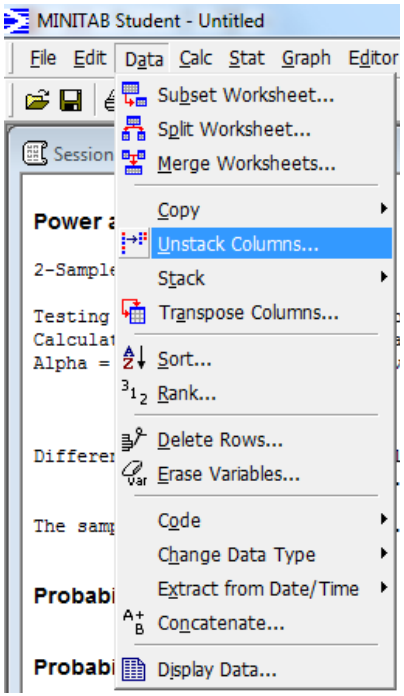
If p-value>0.05, it is normal at 95% confidence, otherwise, it is not.

5. Control charts

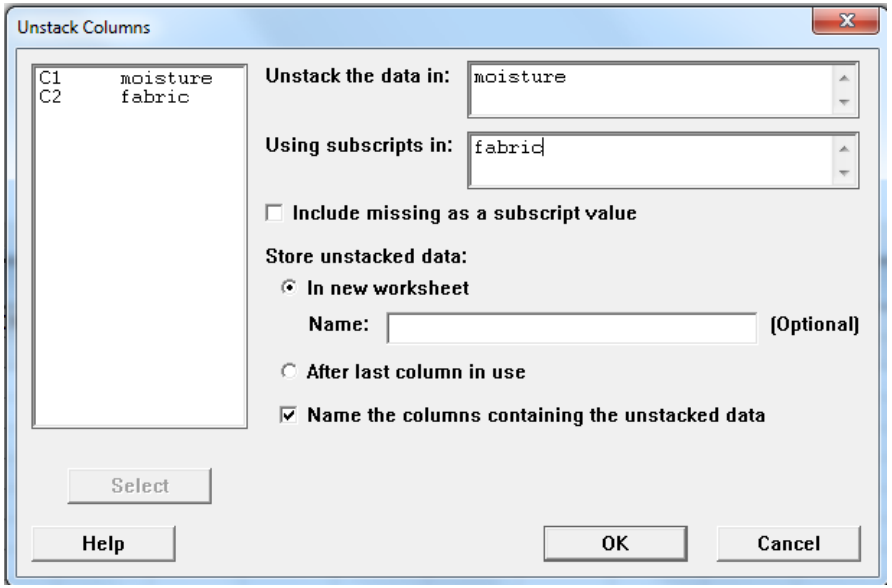
- Devore datasets
- Ex16-08

C1	C2-T
moisture	fabric
12.19999	#1
12.4	#1
12.9	#1
13.2	#1
12.8	#1
13.9	#1
12.2	#1
12.6	#1
14.6	#1
12.8	#1
12.6	#1
13.5	#1
13.4	#1

1. Data-> Unstack Columns



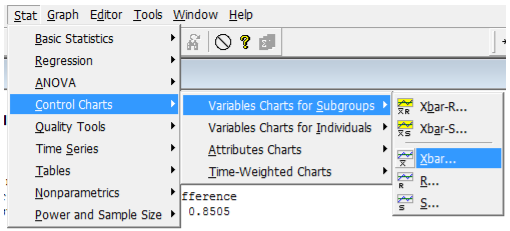
2. Input data and subscripts

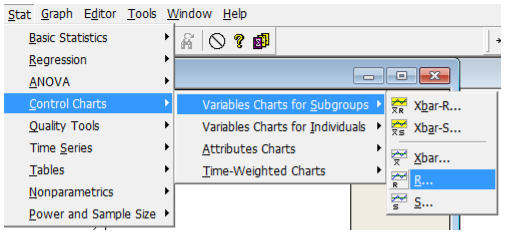


	C1	C2	C3	C4	C5
	moisture_1	moisture_2	moisture_3	moisture_4	moisture_5
1	12.2000	12.1000	13.3000	13.0000	13.0000
2	12.4000	13.3000	12.8000	12.6000	12.9000
3	12.9000	12.7000	14.2000	12.5000	12.9000
4	13.2000	13.0000	13.0000	12.6000	13.9000
5	12.8000	12.3000	12.2000	13.3000	12.0000
6	13.9000	13.4000	13.1000	12.4000	13.2000
7	12.2000	14.4000	12.4000	12.4000	12.5000
8	12.6000	12.8000	13.5000	13.9000	13.1000
9	14.6000	13.4000	12.2000	13.7000	12.5000
10	12.8000	12.3000	12.6000	13.2000	12.8000
11	12.6000	13.1000	12.7000	13.2000	12.3000
12	13.5000	12.3000	12.8000	13.1000	12.9000
13	13.4000	13.3000	12.0000	12.9000	13.1000
14	13.5000	12.4000	13.0000	13.6000	13.4000
15	12.3000	12.8000	13.0000	12.8000	13.5000
16	12.6000	13.4000	12.1000	13.2000	13.3000

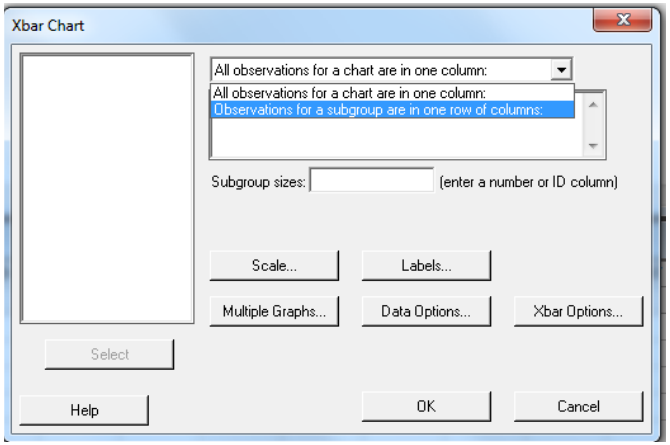
Control Chart

Stat-> Control Charts-> Variables charts for subgroups->Xbar/R

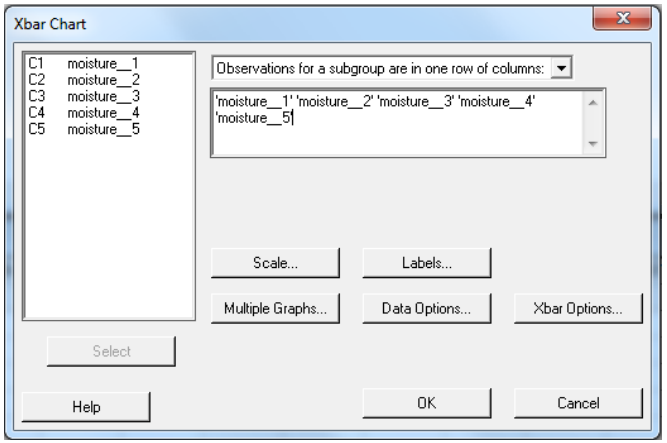


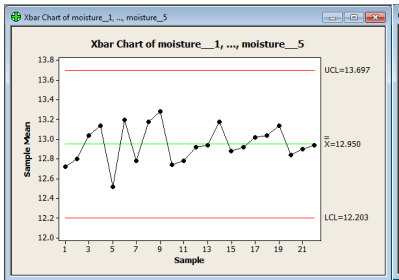


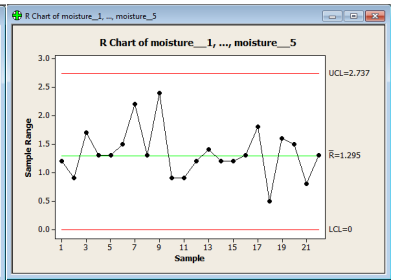
Choose ‘Observations for a subgroup are in one row of columns’



Input all the columns



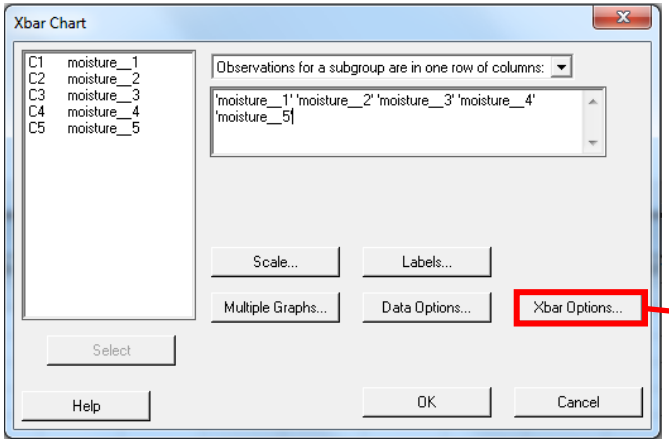




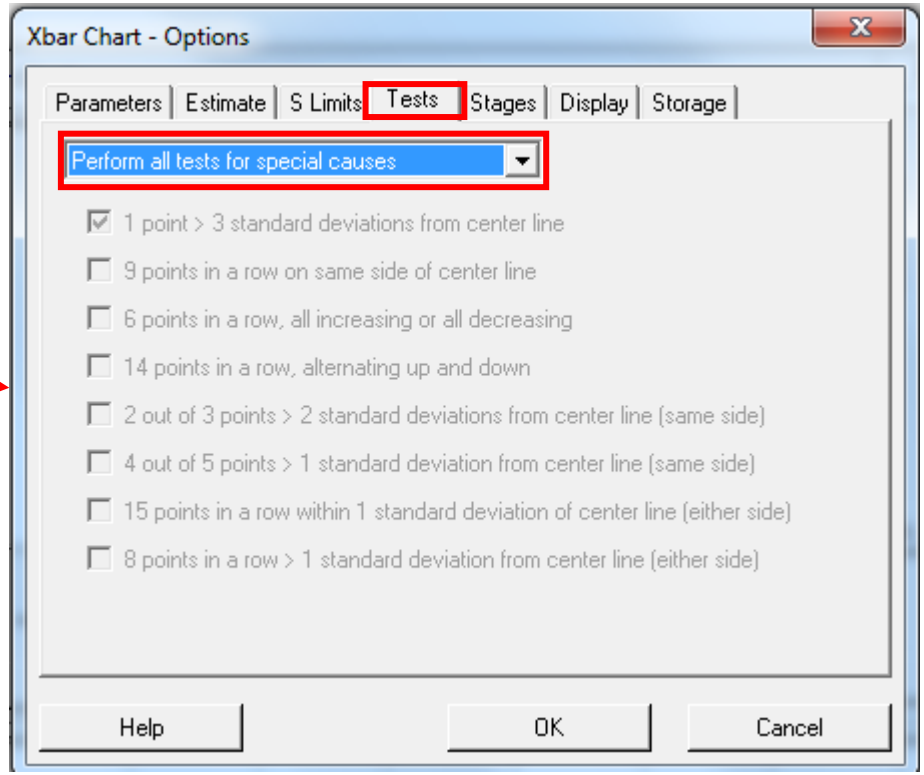
Control chart

- Perform eight tests:

Choose Xbar Options



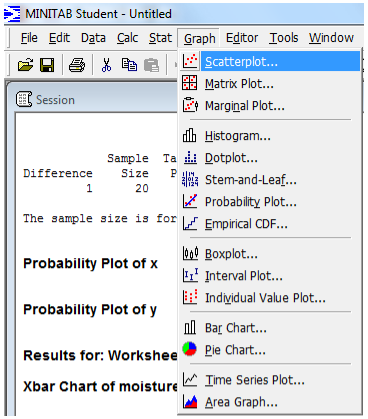
Test-> 'Perform all tests for special causes'

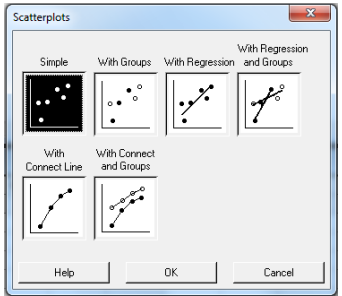


6. Scatter plot, regression

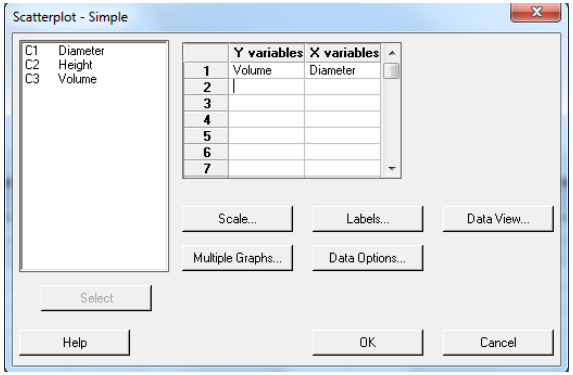
- MINITAB worksheet TREES
- (a) Obtain the scatter plot of the data(volume vs. diameter and volume vs. height)

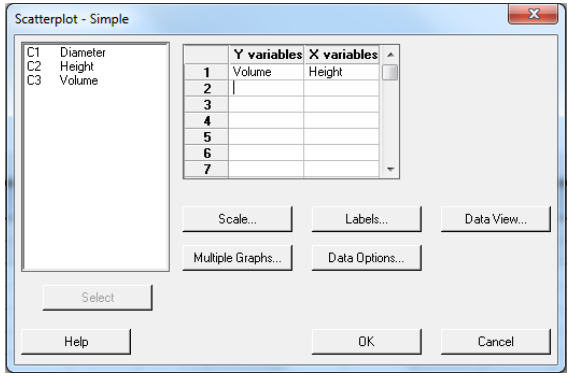
Graph-> Scatterplot

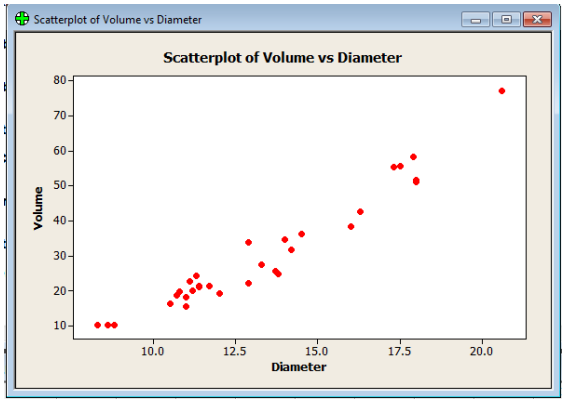


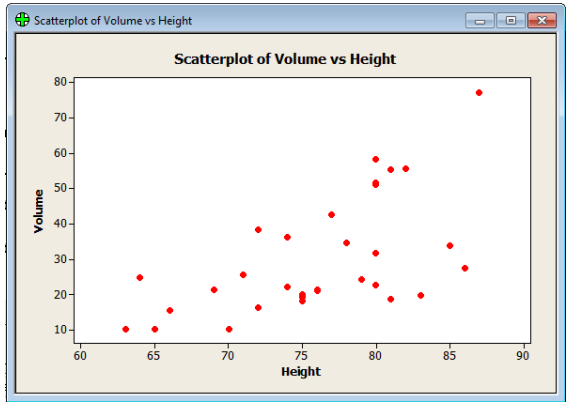


Input Y and X:





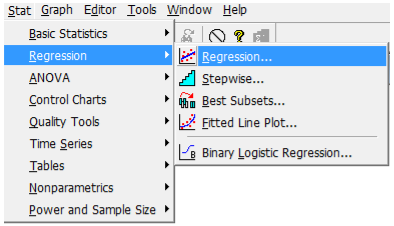




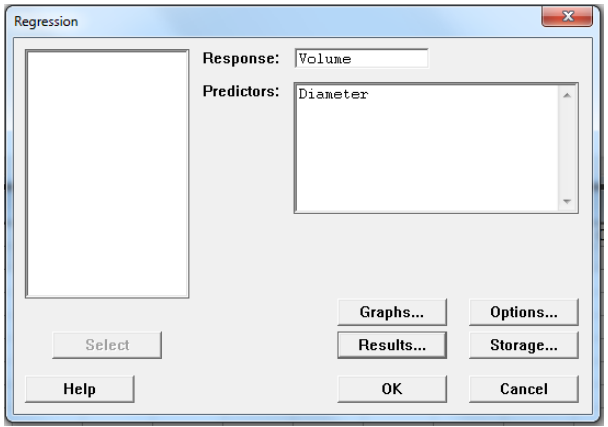
Regression

- (b) Use diameter as a predictor of volume, do a least square analysis and obtain simple linear parameter

Stat-> Regression-> Regression



Input Volume and Diameter



Regression Analysis: Volume versus Diameter

The regression equation is
Volume = - 36.9 + 5.07 Diameter

Predictor	Coef	SE Coef	T	P
Constant	-36.943	3.365	-10.98	0.000
Diameter	5.0659	0.2474	20.48	0.000

S = 4.25199 R-Sq = 93.5% R-Sq(adj) = 93.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	7581.8	7581.8	419.36	0.000
Residual Error	29	524.3	18.1		
Total	30	8106.1			

Unusual Observations

Obs	Diameter	Volume	Fit	SE Fit	Residual	St Resid
31	20.6	77.000	67.413	1.972	9.587	2.55RX

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.