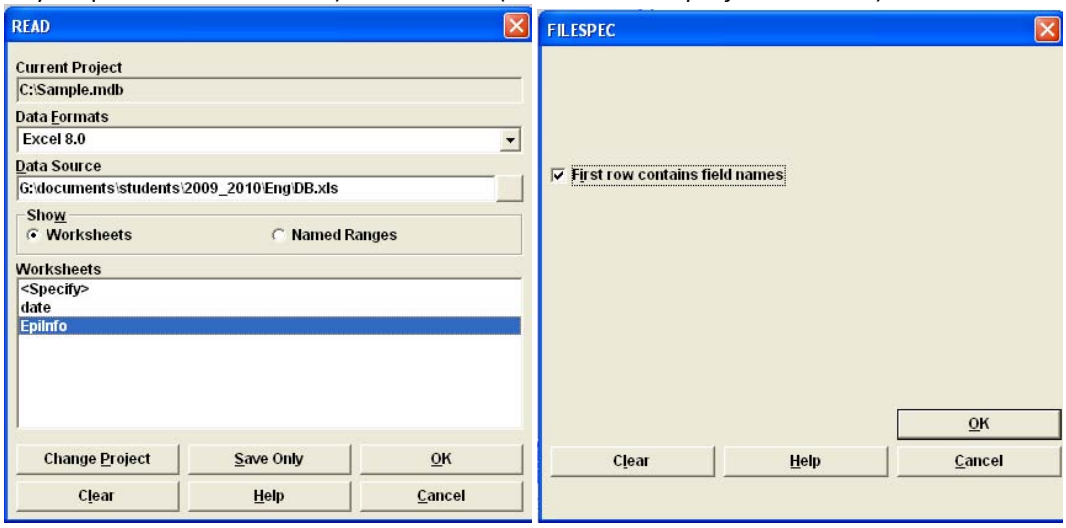
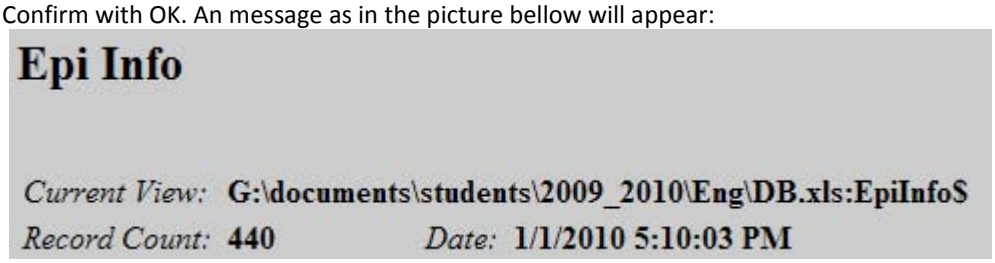
INFERENTIAL STATISTICS IV-INFERENTIAL STATISTICS WITH EXCEL AND EPIINFO: HINTS

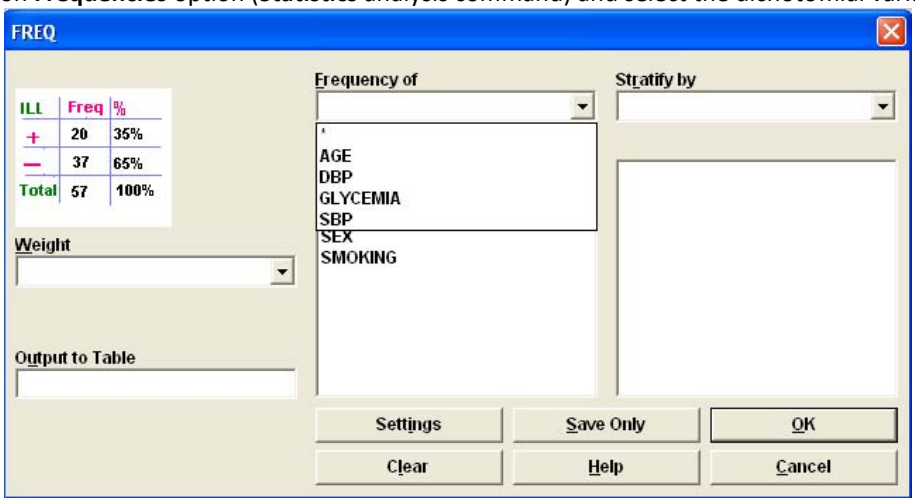
1. Save the Excel file into your partition on the server on a folder corresponding to the number of practical activity. Use IF predefine function, Microsoft Excel.
2. Use [Edit - Past Special (values and number formats)]. Save the document and close it.
3. Open Analysis module of EpiInfo ([Start – All Programs – Epi Info - Analysis]). Activate Read(Import) option.
    a. Define the type of document which you wand to open (Excel file) & data source (find the DB.xls file on your partition on the server) & Worksheet (select here the *EpiInfo* work sheet):



    b. Confirm with OK. An message as in the picture bellow will appear:

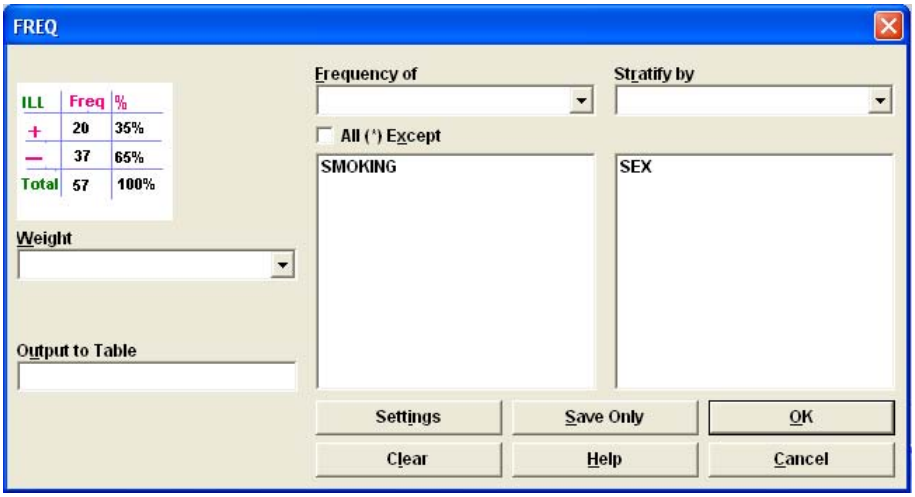


    c. Verify if the open data are the correct ones. From **Statistics** command activates with double click the **List** option. Chose all (*) and validate with OK. It is correct if you have the HT and Diabetes status for all patients (as yes/no variables).
4. Double click on **Frequencies** option (**Statistics** analysis command) and select the dichotomial variables:



validate the selections with OK. You can obtain the frequencies tables for all variables using the **Frequencies** command just one time.

5. Create a stratified frequency table [Frequencies – Frequency of SMOKING – Stratified by SEX]:



To answer the question you will look at percent of smoking = yes for SEX= F and SEX=M and to associated 95% Conf Limits:

SMOKING, SEX=F

Forward

SMOKING	Frequency	Percent	Cum Percent
1	31	14.8%	14.8%
2	179	85.2%	100.0%
Total	210	100.0%	100.0%

95% Conf Limits

1 10.3% 20.3%

2 79.7% 89.7%

SMOKING, SEX=M

Back Forward Current Procedure

SMOKING	Frequency	Percent	Cum Percent
1	72	31.3%	31.3%
2	158	68.7%	100.0%
Total	230	100.0%	100.0%

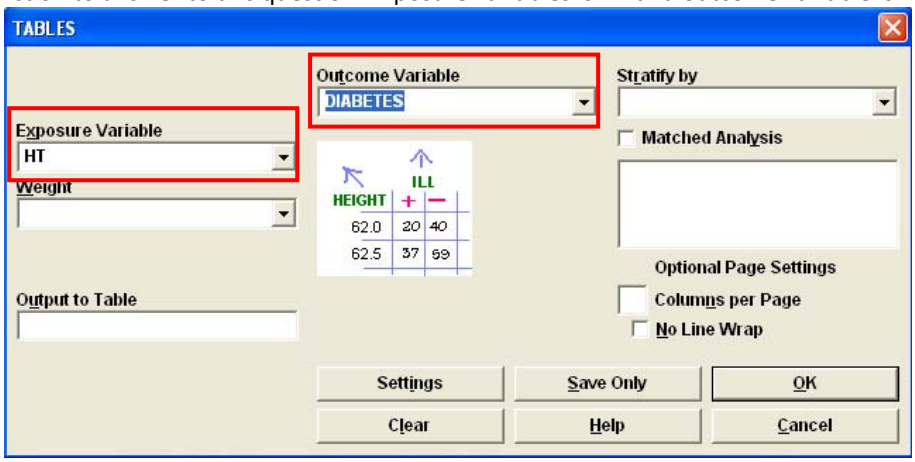
95% Conf Limits

1 25.4% 37.7%

2 62.3% 74.6%

Since the 95% confidence limits of percentage are not overlap on each other we can state that the percent of smoking male is significantly higher than the percent of smoking female.

6. Similar with hints for request 5.
7. Similar with hints for request 5.
8. Use **Table** function to answer to this question. Exposure variables is HT and outcome variable is DIABETES



The following results will be obtained:

Single Table Analysis

	Point Estimate	95% Confidence Interval Lower	Upper	
PARAMETERS: Odds-based				
Odds Ratio (cross product)	0.6740	0.4455	1.0199	(T)
Odds Ratio (MLE)	0.6746	0.4449	1.0215	(M)
		0.4360	1.0421	(F)
PARAMETERS: Risk-based				
Risk Ratio (RR)	0.7562	0.5641	1.0137	(T)
Risk Difference (RD%)	-8.1267	-16.6733	0.4198	(T)

(T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact)

STATISTICAL TESTS	Chi-square	1-tailed p	2-tailed p
Chi-square - uncorrected	3.5028		0.0612660794
Chi-square - Mantel-Haenszel	3.4948		0.0615613038
Chi-square - corrected (Yates)	3.1182		0.0774238886
Mid-p exact		0.0314975454	
Fisher exact		0.0388636513	

Since the question is about risk factors, we will look and interpret Risk Ration or Odds Ratio (depending on research methodology – Second semester) as parameters and associated confidence interval.

Since the values (both in this case OR = Odds Ratio and RR = Risk Ratio) are not higher than 1 the HT could not be considered a risk factor for Diabetes.

9. A. Similar with hints gave for request 8. B. Do not forget to stratify the results on SEX.
10. Similar with hints gave for request 8 but this time we will interpret the Chi-Square test.
    a. The observed contingency table is:

DIABETES

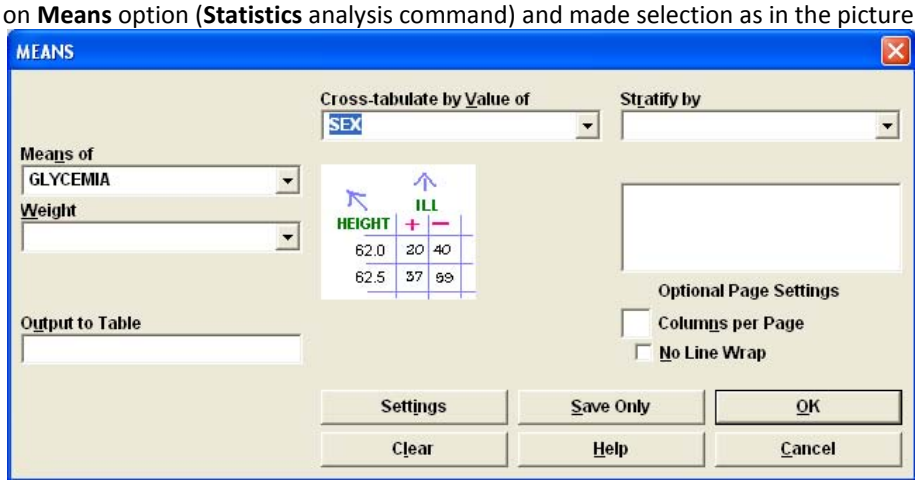
OBESITY	1	2	TOTAL
1	41	63	104
Row %	39.4	60.6	100.0
Col %	32.3	20.1	23.6
2	86	250	336
Row %	25.6	74.4	100.0
Col %	67.7	79.9	76.4
TOTAL	127	313	440
Row %	28.9	71.1	100.0
Col %	100.0	100.0	100.0

    b. Since all observed values are higher than 5 you will look on Chi-Square uncorrected or Mantel-Haenszel:

STATISTICAL TESTS	Chi-square	1-tailed p	2-tailed p
Chi-square - uncorrected	7.3958		0.0065387529
Chi-square - Mantel-Haenszel	7.3790		0.0066001352
Chi-square - corrected (Yates)	6.7377		0.0094408712
Mid-p exact		0.0039658922	
Fisher exact		0.0053218909	

Since The values of statistical tests are higher than critical values for 1 degree of freedom and the p values are less than 5% it can be concluded that obesity and diabetes are dependent.

11. Double click on **Means** option (**Statistics** analysis command) and made selection as in the picture bellow:



a. Descriptive statistics:

Descriptive Statistics for Each Value of Crosstab Variable

	Obs	Total	Mean	Variance	Std Dev
F	210	22395.0000	106.6429	1418.3838	37.6614
M	230	28056.0000	121.9826	1943.5892	44.0862

	Minimum	25%	Median	75%	Maximum	Mode
F	62.0000	83.0000	93.0000	107.0000	210.0000	83.0000
M	51.0000	85.0000	101.0000	174.0000	199.0000	82.0000

It could be observed that the mean of Glycaemia is higher in Male compared with Female since the modes are almost equal. The same observation is also correct for variance.

b. Inferential statistics: the results of comparisons are presented as in the picture bellow:

ANOVA, a Parametric Test for Inequality of Population Means

(For normally distributed data only)

Variation	SS	df	MS	F statistic
Between	25830.3985	1	25830.3985	15.2574
Within	741524.1447	438	1692.9775	
Total	767354.5432	439		

T Statistic = 3.9061

P-value = 0.0001

Bartlett's Test for Inequality of Population Variances

Bartlett's chi square= 5.3617 df=1 P value=0.0206

A small p-value (e.g., less than 0.05 suggests that the variances are not homogeneous and that the ANOVA may not be appropriate.

Mann-Whitney/Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

Kruskal-Wallis H (equivalent to Chi square) = 13.0846
Degrees of freedom = 1
P value = 0.0003

As is stated in the picture, the ANOVA test is appropriate if the data are normally distributed and homogenous. Since the data are not homogenous (Bartlett chi-square = 5.3617, p = 0.0206) the Kruskals-Wallis test for two groups is the appropriate test. The result of this test revealed that the mean of glycaemia in male is significantly higher compared with the mean of glycaemia in female.

12. Your presentation must have the following structure:
    a. Title: "Analysis of Risks Factors for Diabetes" & Author (your full name) & Author affiliation (University - Faculty) (First page / slide)
    b. Content (Second page / slide). If use Microsoft Word use Styles and create the Content automatically. If use Microsoft PowerPoint include here the title of each slides in the presentation.
    c. On a new page / slide provide the answer for request from 5 to 12. Include the following information: results and interpretation.
    d. Save the file in a folder dedicated to the present practical activity with name Obesity Risk Factors Analysis.