Lab 8: Comparing proportions—Fisher's Exact Test and Chi-squared test

Introduction and Objective:

In the last lecture, we discussed how to compare proportions using Fisher's exact test and Chi-squared test. Also, we discussed under which circumstances, we should choose which test. We learned how to compute the P values by hand using these two tests. In today's lab, we are going to learn how to use R to perform Fisher's exact test and Chi-squared test. At the end of the lab, I would like to introduce how to use another computer program, Prism/Graph Pad, to perform such tests.

There is an in-class assignment due at the end of the lab.

The following files you should have received and are needed for today's class:

1) AntiCancerDrugContingencyTable1.csv
2) ThreeDrugTreatments1.csv
3) TwoDrugsLargeNumbers.csv
4) TwoDrugsSmallNumbers.csv
5) DrugScreening1.csv
6) Mendel_F2_Phenotypic_Ratios.csv

All the R command lines are in **bold**; all the notes are following a #; all the R results directly follow the R codes/command lines and are not in bold or following a #.

I. Use Fisher's exact test and Chi-squared test in R to compare proportions

```
setwd("C:/R")

# Let's use an example we used in lecture:

a<-read.csv("AntiCancerDrugContingencyTable1.csv")

View(a)

dim(a)

[1] 3 4
```

```
b<-data.frame(a[1:2,1:3])

b

        Treatments Metastasized No.Metastasis
1          Placebo          427             2
2 Anti-cancer drug           85           345

dim(b)

[1] 2 3

    # We select the numbers only and deposit them into an object named as "c":

c<-b[,2:3]

fisher.test(c)

    # The R built-in function "fisher.test" allows you to perform Fisher's exact test on the numbers in a contingency table only.

Fisher's Exact Test for Count Data

data:  c

p-value < 2.2e-16

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

  226.1546 8192.0000

sample estimates:

odds ratio

  850.0433

   # From the result you can see, it tells you the p-value from Fisher's exact test. Also, it tells you, if based on the p-value, the null hypothesis (H0) would be rejected, what the alternative hypothesis (HA) would be adopted.
```

# We can also try performing a Chi-squared test on the same data by using the built-in R function “chisq.test”:

```
chisq.test(c)
```

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  c

X-squared = 564.183, df = 1, p-value < 2.2e-16
```

# We can see both tests gave you a similar result. However, as we talked about this in class, under such circumstances (when there are only two categories), we prefer to use Fisher’s exact test.

# Now, let’s take a look at another example:

```
TD<-read.csv("ThreeDrugTreatments1.csv")

View(TD)

dim(TD)
```

```
[1] 3 4
```

# We select the numbers only and then perform Fisher’s exact test:

```
TD1<-TD[,2:3]

fisher.test(TD1)
```

```
Error in fisher.test(TD1) : FEXACT error 6.

LDKEY is too small for this problem.

Try increasing the size of the workspace.
```

# As you can see, when you tried to use Fisher's exact test on 3 categories (as TD1 has 3 rows), it didn’t work and R gave you an error message. By checking the dimensions of TD1, we can see there are 3 rows. So, instead of using Fisher’s exact test, we can choose to perform a Chi-squared test on this data (TD1).

```
dim(TD1)
```

```
[1] 3 2
```

```
chisq.test(TD1)


        Pearson's Chi-squared test


data:  TD1

X-squared = 1363.491, df = 2, p-value < 2.2e-16
```

```
        # We can see, when more categories and contingencies present, the calculations of Fisher's
exact test tremendously increase, so even it is hard to be handled by the computer. In this case,
we use the Chi-squared test to generate an approximate P value.

        # Now, we look at a third example:
```

```
TL<-read.csv("TwoDrugsLargeNumbers.csv")

View(TL)
```

```
        # Challenge Question 1: After you looked at this table, what type of analyses would you
perform to compare the proportions, Fisher's exact test or Chi-squared test?
```

```
dim(TL)

[1] 2 4

TL1<-TL[,2:3]

fisher.test(TL1)

        Fisher's Exact Test for Count Data

data:  TL1

p-value < 2.2e-16

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 6.162527 7.661716

sample estimates:

odds ratio
```

```
  6.868522

chisq.test(TL1)

        Pearson's Chi-squared test with Yates' continuity correction

data:  TL1

X-squared = 1355.632, df = 1, p-value < 2.2e-16

         # As you can see, even with large numbers, when there are only 2 contingencies, R can handle it with ease using Fisher’s exact test.

         # Challenge Question 2: What about data with small numbers? Which test is more appropriate?

TS<-read.csv("TwoDrugsSmallNumbers.csv")

View(TS)

dim(TS)

[1] 2 3

TS1<-TS[,2:3]

fisher.test(TS1)

        Fisher's Exact Test for Count Data

data:  TS1

p-value = 0.4875

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 0.0793999 3.0322824

sample estimates:

odds ratio

 0.5563313
```

```
chisq.test(TS1)
```

```
	Pearson's Chi-squared test with Yates' continuity correction

data:  TS1
X-squared = 0.1702, df = 1, p-value = 0.68
Warning message:
In chisq.test(TS1) : Chi-squared approximation may be incorrect
```

# As you can see from the above when the numbers are small, both Fisher’s exact test and Chi-squared test can handle the computation and calculate a p-value. However, when both methods are available, we prefer to use Fisher’s exact test, which will calculate a much more accurate p-value than the Chi-squared test.

# Instead let R to determine ratios/proportions for the Chi-squared test, what if you have your proposed ratios from your theory? Like what we discussed in class:

# If Mendel’s theory, which later was further developed as the independent segregation of the alleles, is correct, the phenotypic ratio of the F2 peas would be 9:3:3:1. However, the ratio Mendel observed was 315:108:101:32, which mathematically does not equal to 9:3:3:1. But whether or not the difference between his observation and the proposed 9:3:3:1 ratio is not statistically significant and thus his observation fits the model? Whether his observation supports his theory? To answer such kind questions, we need to perform Chi-squared test (goodness-of-fit test). How can we perform the Chi-squared test using R to solve this problem?

```
MendelData<-read.csv("Mendel_F2_Phenotypic_Ratios.csv")
```

# Once you deposit the data in object “MendelData”, you need to select originally observed data for the Chi-squared test, and also inform R what the proposed proportions/ratios (given probabilities) are by using the argument “p=”:

```
chisq.test(MendelData[1:4,2],p=MendelData[1:4,4])
```

```
	Chi-squared test for given probabilities

data:  MendelData[1:4, 2]

X-squared = 0.47002, df = 3, p-value = 0.9254
```

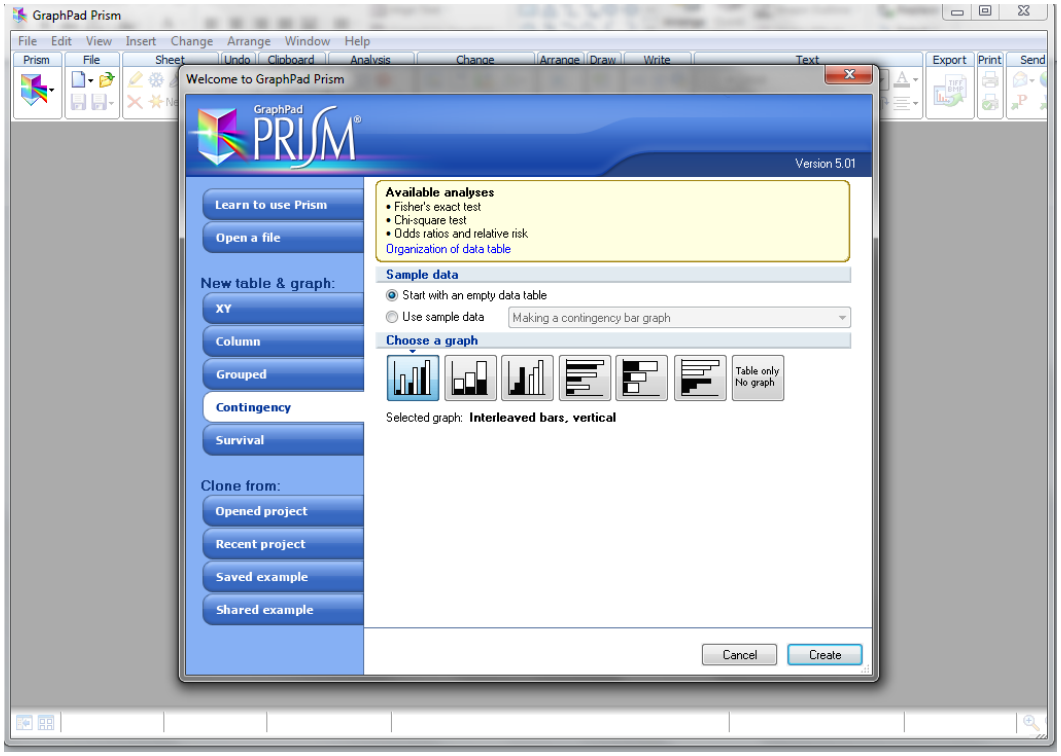
# You can compare what we calculated manually in class with the results you got by using R and see they are the same.

II. Use Prism/GraphPad to perform Fisher's exact test and Chi-squared test:

# On the left side, select the tab for "contingency" table

# Choose the grouped bar chart (the first one under "Choose a graph")

# Click "create"



Available analyses
- Fisher's exact test
- Chi-square test
- Odds ratios and relative risk

Organization of data table

Sample data

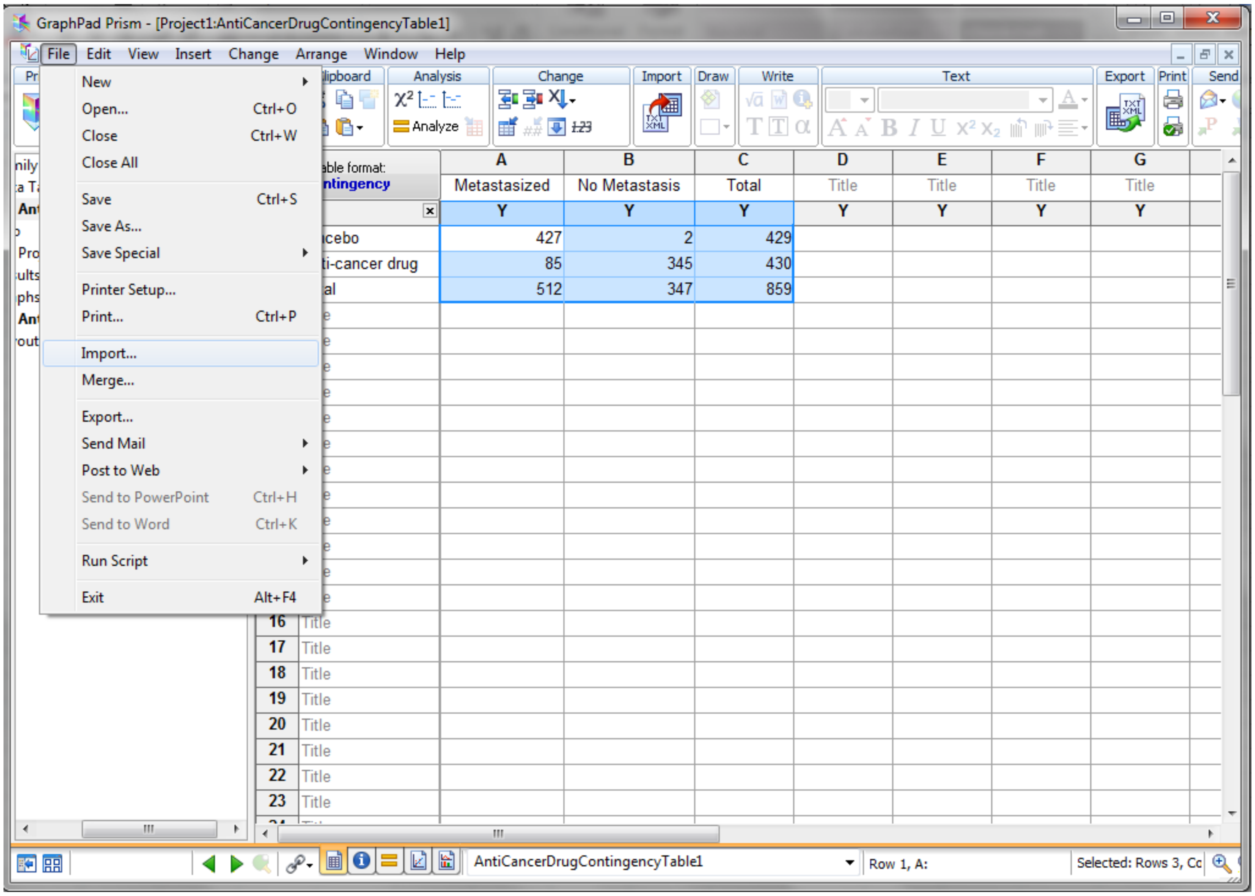
● Start with an empty data table

○ Use sample data   Making a contingency bar graph

Choose a graph

Selected graph: **Interleaved bars, vertical**

# You may type in your data, or you may choose to import the data. If you chose to import the data, the screenshot you got would look like the following:



	A	B	C	D	E	F	G
	Metastasized	No Metastasis	Total	Title	Title	Title	Title
	Y	Y	Y	Y	Y	Y	Y
cebo	427	2	429				
ti-cancer drug	85	345	430				
al	512	347	859				

# Then click on “Analyze”:

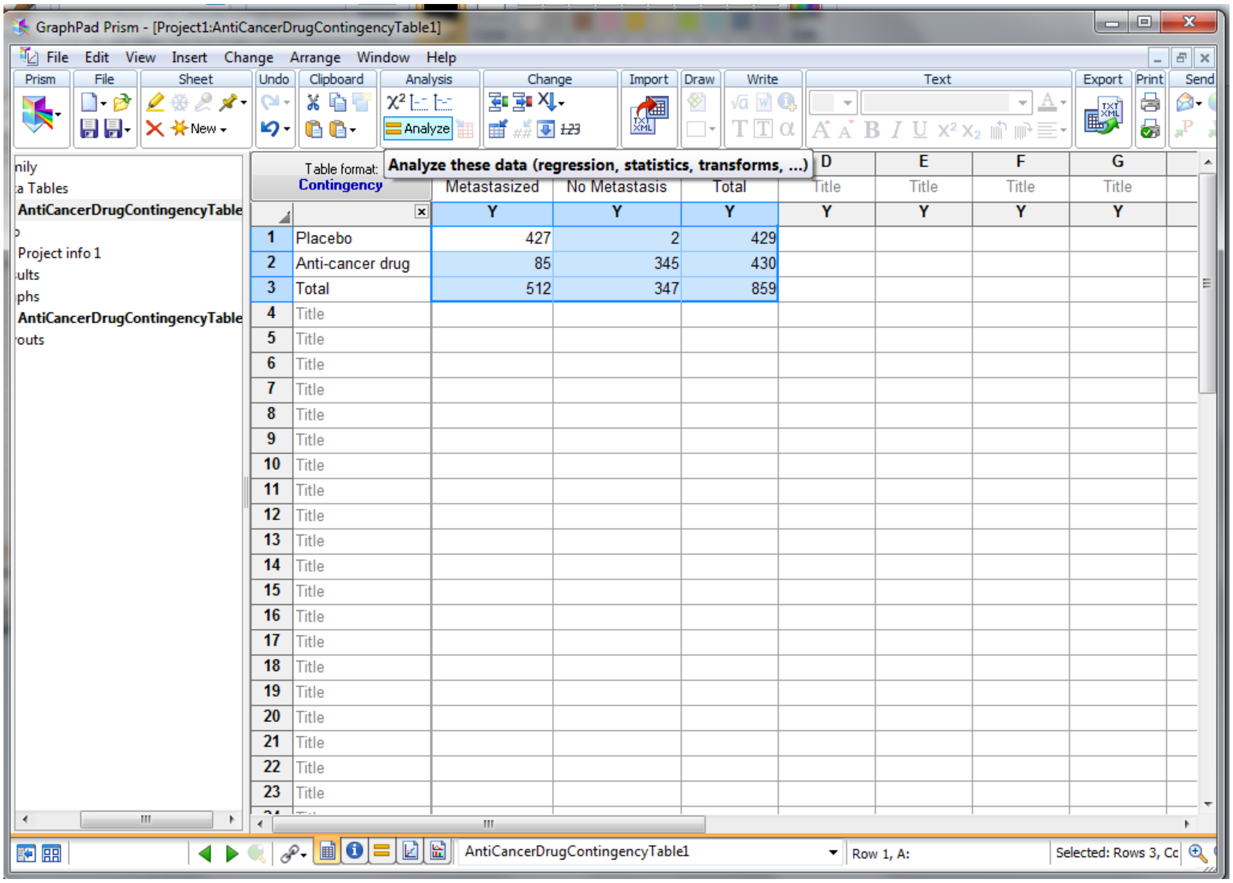
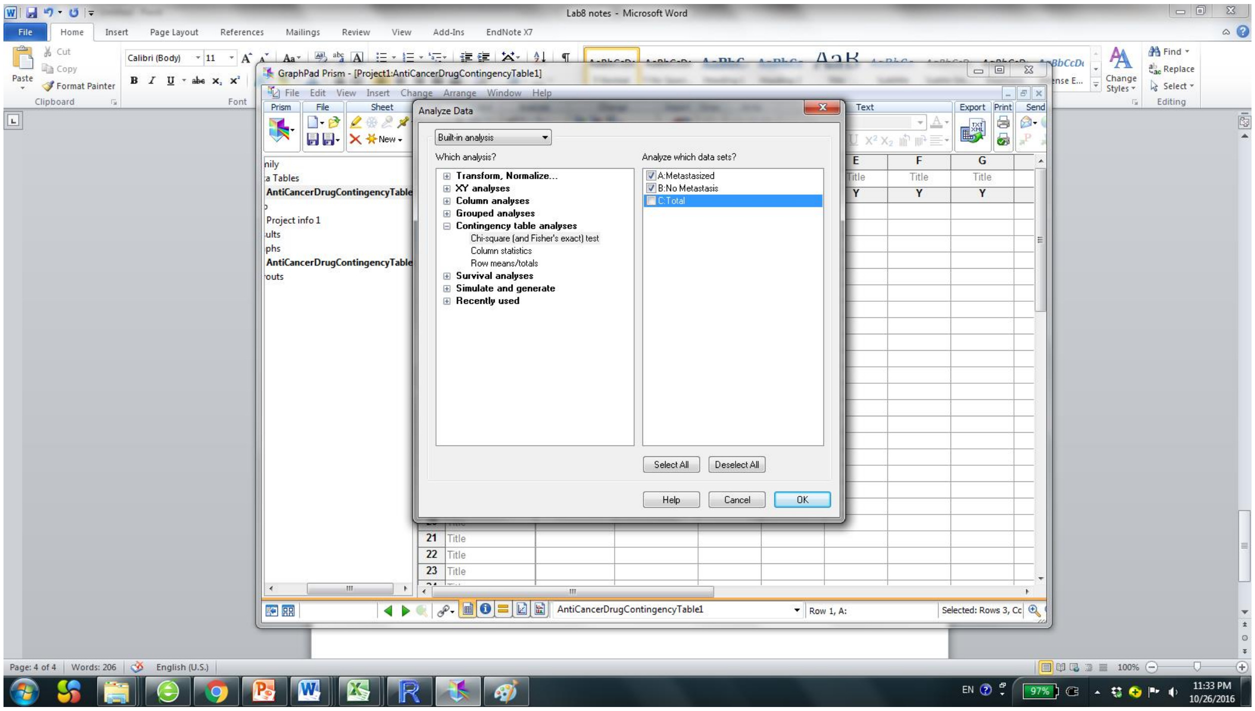

GraphPad Prism - [Project1:AntiCancerDrugContingencyTable1]
File Edit View Insert Change Arrange Window Help
Prism File Sheet Undo Clipboard Analysis Change Import Draw Write Text Export Print Send
Analyze
Analyze these data (regression, statistics, transforms, ...)
AntiCancerDrugContingencyTable1
Row 1, A:
Selected: Rows 3, Cc

Table format: Contingency

		Metastasized	No Metastasis	Total
		Y	Y	Y
1	Placebo	427	2	429
2	Anti-cancer drug	85	345	430
3	Total	512	347	859

# Uncheck “Total”:


Lab8 notes - Microsoft Word
GraphPad Prism - [Project1:AntiCancerDrugContingencyTable1]
Analyze Data
Select All
Deselect All
Help
Cancel
OK

Built-in analysis

Which analysis?

- Transform, Normalize...
- XY analyses
- Column analyses
- Grouped analyses
- Contingency table analyses
  - Chi-square (and Fisher's exact) test
  - Column statistics
  - Row means/totals
- Survival analyses
- Simulate and generate
- Recently used

Analyze which data sets?

☑ A:Metastasized
☑ B:No Metastasis
□ C:Total

# Actually, after you import the data, you can delete the row and the column for the grand total and the marginal totals.

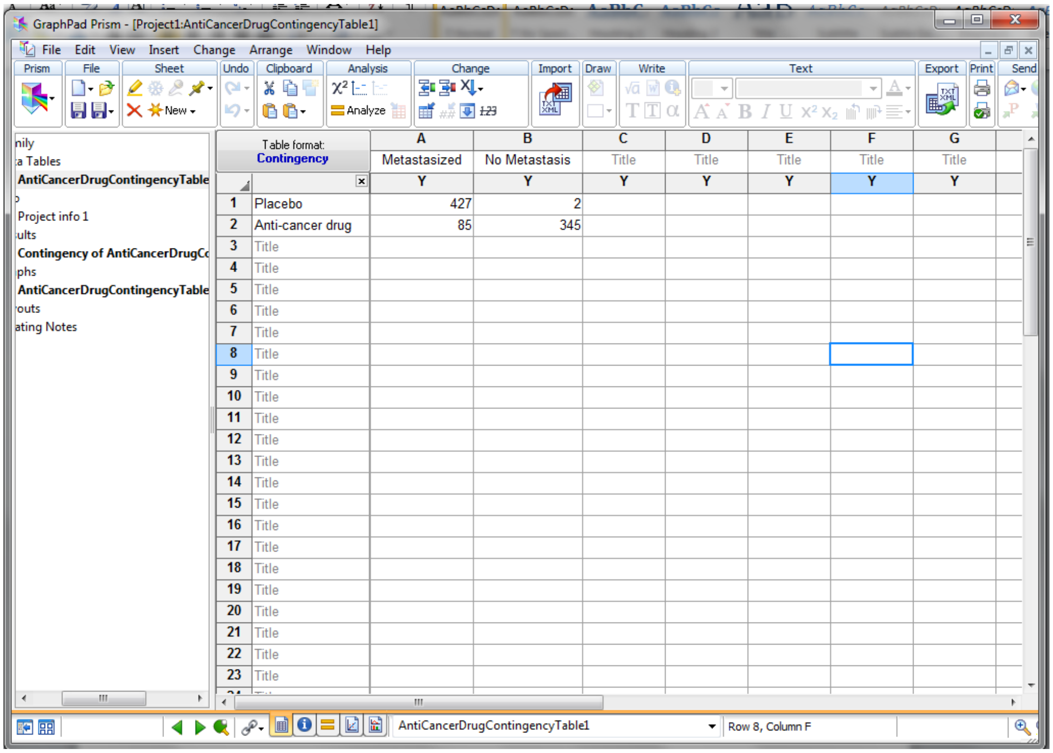
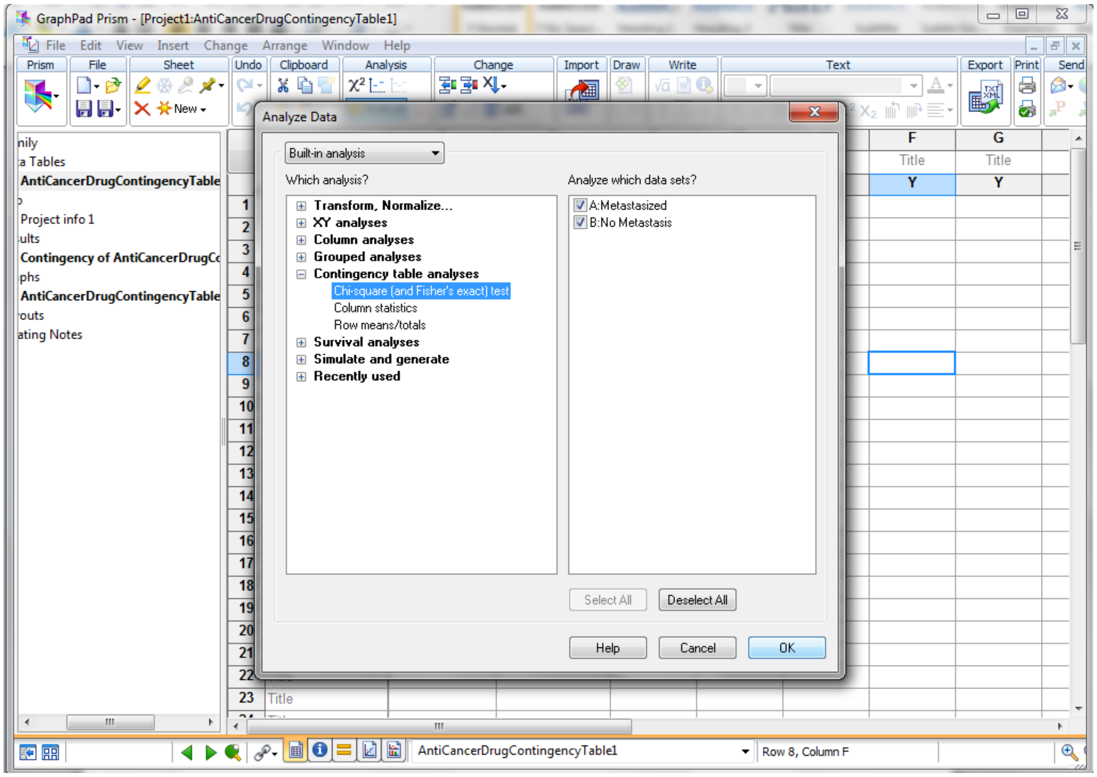

GraphPad Prism - [Project1:AntiCancerDrugContingencyTable1]
File Edit View Insert Change Arrange Window Help
Prism File Sheet Undo Clipboard Analysis Change Import Draw Write Text Export Print Send
New
Analyze
nily
a Tables
AntiCancerDrugContingencyTable
Project info 1
ults
Contingency of AntiCancerDrugCo
phs
AntiCancerDrugContingencyTable
outs
ating Notes
AntiCancerDrugContingencyTable1
Row 8, Column F

Table format: Contingency	A	B	C	D	E	F	G
	Metastasized	No Metastasis	Title	Title	Title	Title	Title
	Y	Y	Y	Y	Y	Y	Y
1 Placebo	427	2					
2 Anti-cancer drug	85	345					

# Click on the "Analyze":


GraphPad Prism - [Project1:AntiCancerDrugContingencyTable1]
File Edit View Insert Change Arrange Window Help
Prism File Sheet Undo Clipboard Analysis Change Import Draw Write Text Export Print Send
New
nily
a Tables
AntiCancerDrugContingencyTable
Project info 1
ults
Contingency of AntiCancerDrugCo
phs
AntiCancerDrugContingencyTable
outs
ating Notes
AntiCancerDrugContingencyTable1
Row 8, Column F

Analyze Data

Built-in analysis

Which analysis?

- ⊞ Transform, Normalize…
- ⊞ XY analyses
- ⊞ Column analyses
- ⊞ Grouped analyses
- ⊟ Contingency table analyses
  - Chi-square (and Fisher's exact) test
  - Column statistics
  - Row means/totals
- ⊞ Survival analyses
- ⊞ Simulate and generate
- ⊞ Recently used

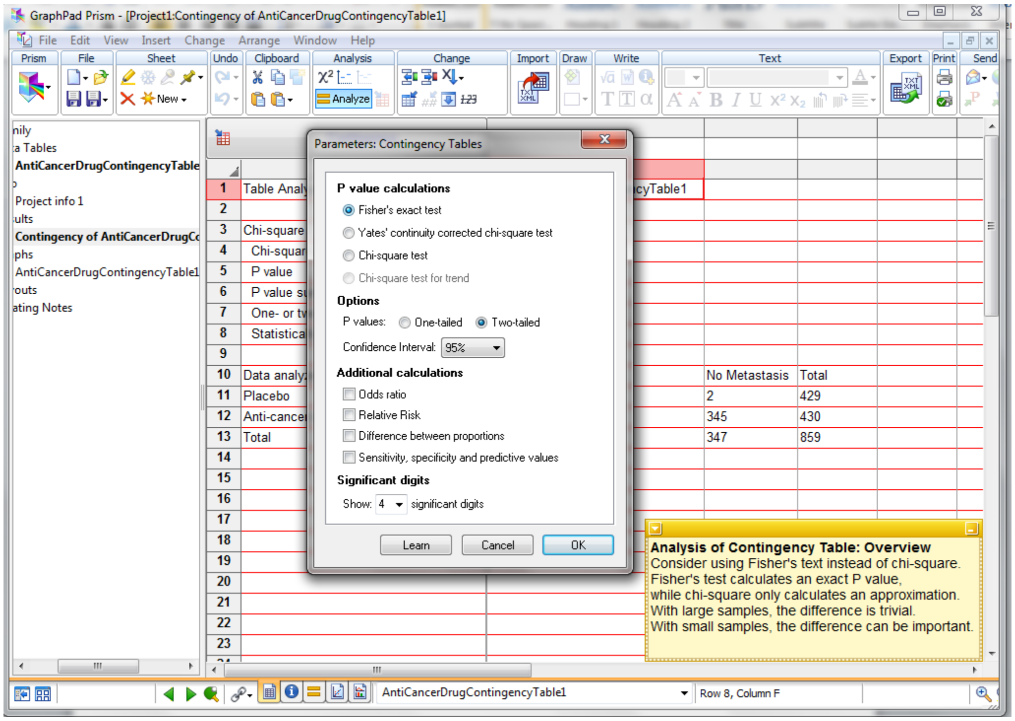
Analyze which data sets?

☑ A:Metastasized
☑ B:No Metastasis

Select All | Deselect All

Help | Cancel | OK

# Under "Contingency table analyses", choose "Chi-square (and Fisher's exact) test, make sure both "Metastasized" and "No Metastasis" are checked, and then click "OK". Next, choose "Fisher's exact test":


GraphPad Prism - [Project1:Contingency of AntiCancerDrugContingencyTable1]
File Edit View Insert Change Arrange Window Help
Prism File Sheet Undo Clipboard Analysis Change Import Draw Write Text Export Print Send
AntiCancerDrugContingencyTable1
Row 8, Column F

Parameters: Contingency Tables

P value calculations

- ● Fisher's exact test
- ○ Yates' continuity corrected chi-square test
- ○ Chi-square test
- ○ Chi-square test for trend

Options

P values: ○ One-tailed ● Two-tailed

Confidence Interval: 95%

Additional calculations

- □ Odds ratio
- □ Relative Risk
- □ Difference between proportions
- □ Sensitivity, specificity and predictive values

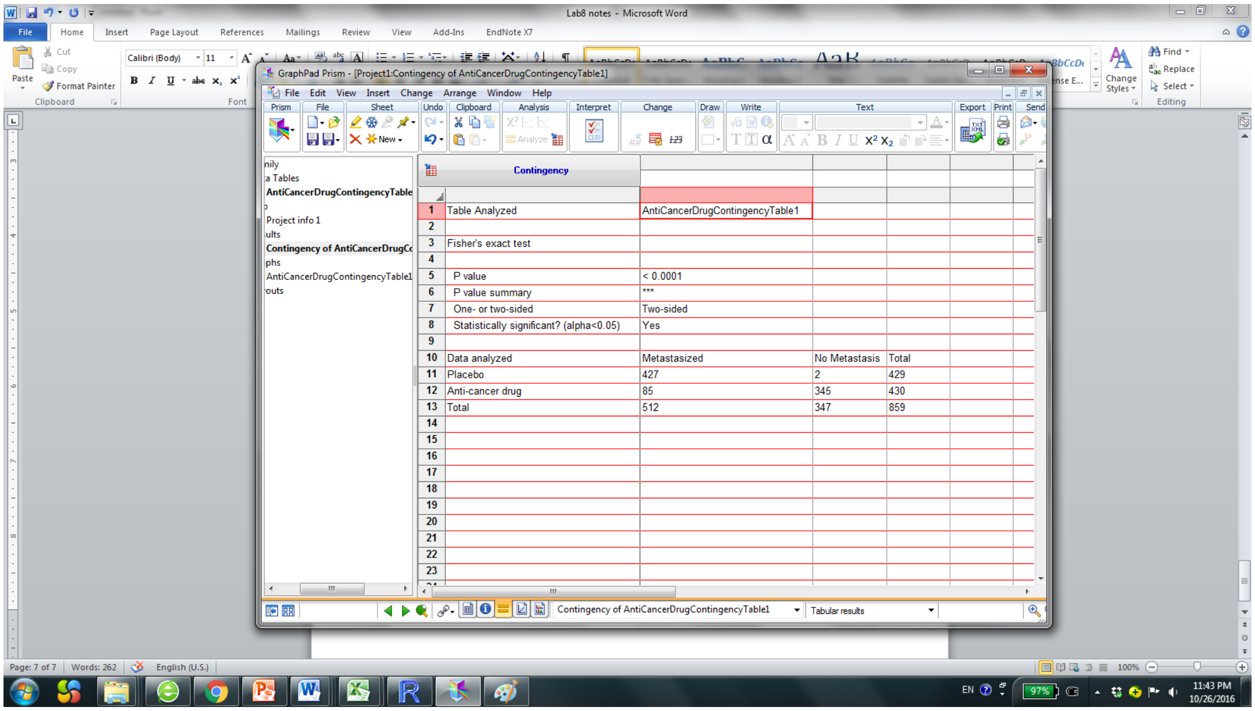
Significant digits

Show: 4 significant digits

Learn | Cancel | OK

Analysis of Contingency Table: Overview
Consider using Fisher's text instead of chi-square.
Fisher's test calculates an exact P value,
while chi-square only calculates an approximation.
With large samples, the difference is trivial.
With small samples, the difference can be important.

# Once you click "OK", you will get:


Lab8 notes - Microsoft Word
GraphPad Prism - [Project1:Contingency of AntiCancerDrugContingencyTable1]
File Edit View Insert Change Arrange Window Help
Contingency of AntiCancerDrugContingencyTable1
Tabular results
Page: 7 of 7 Words: 262 English (U.S.)
11:43 PM
10/26/2016

	Contingency			
1	Table Analyzed	AntiCancerDrugContingencyTable1		
2				
3	Fisher's exact test			
4				
5	P value	< 0.0001		
6	P value summary	***		
7	One- or two-sided	Two-sided		
8	Statistically significant? (alpha<0.05)	Yes		
9				
10	Data analyzed	Metastasized	No Metastasis	Total
11	Placebo	427	2	429
12	Anti-cancer drug	85	345	430
13	Total	512	347	859

Answers to the challenge questions:

Challenge Question 1:

Of course, a Chi-squared test can handle the analysis. And because there are only two categories for the contingencies, a Fisher's exact test also can be performed. When both Fisher's exact test and Chi-squared test can be used, we prefer to use Fisher's exact test as it generates an exact p-value.

Challenge Question 2:

Fisher's exact test is more appropriate for handling small numbers. The p-value from a Fisher's exact test is more accurate than from a Chi-squared test, in particular for data with small numbers.

Group Assignment 6:

Assume you did a clinical trial III testing two drugs for their anti-cancer effects. The result is summarized in the table "DrugScreening1.csv".

1) Generate a data frame and name it as "AB". This "AB" includes the data of reduced tumor size and increased tumor size for the two tested drugs only. Write R codes to rename the column names as "Reduced Tumor Size" and "Increased Tumor Size". Show your R commands and the result. (2pts)

2) For tumors having reduced size, and for the tumors remaining the same size, compare Drug1 with the placebo. Compute a P value, to see if Drug1 is better than the placebo, which means if the ratio of the number of tumors having reduced size to the number of tumors remaining the same size is smaller for Drug1than for the placebo. What's your H0? State your own alpha. Which test should you use? Show your R commands and the result. (4pts)

3) For tumors having reduced size and for the tumors increased size, compare Drug2 with Drug1. Is Drug1 better than Drug2? In other word, if the ratio of the number of tumors

having reduced size to the number of tumors increased size is smaller for Drug1than for Drug2? Show your R commands and your conclusions. (2pts)

4) Compare Drug 1 with the placebo for tumors having reduced size and tumors having increased size. Is Drug 1 better than the placebo? Show your R commands and your conclusions. (2pts)