

WOOOHOOO!

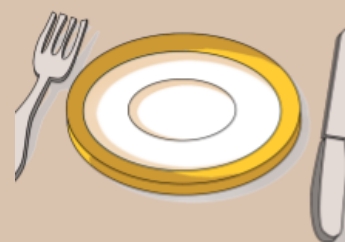
The SMART Board is up :-)

Get:

- 1 half sheet of paper
- supply bin

On Your Desk:

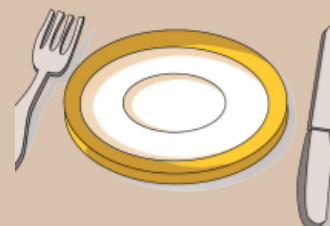
- notebook
- calculator



Sep 22-7:26 AM

3.1

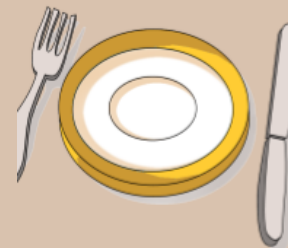
Contingency Tables



Sep 22-7:26 AM

Does smoking cause a shorter life span?

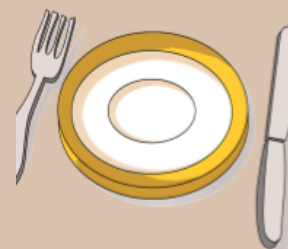
- One study in the 70's found that 24% of the smokers had passed away 20 years later, and 31% of the non-smokers had passed away!
- How can this be?? Is there something else that could explain this...?



Sep 22-7:26 AM

We want to compare variables...

- For instance, is there a relationship between the amount of gasoline used by cars and air pollution?
- Do schools with a higher per-student funding tend to have higher SAT averages?
- Does a glass of wine a day increase life span?
- Does an apple a day really keep the doctor away?



Sep 22-7:26 AM

Explanatory variable

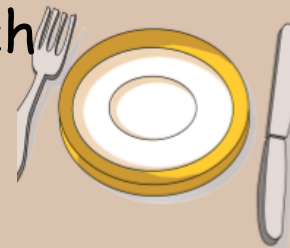
(Independent variable)

defines the groups to be compared with respect to values on the response variable

Response variable

(Dependent Variable)

the outcome variable on which comparisons are made



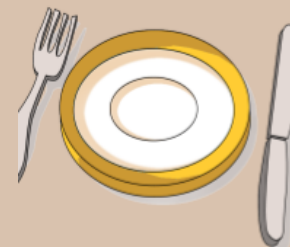
Sep 22-7:26 AM

Example: Explanatory/Response

of drinks consumed/Blood alcohol level

Amount of study time/Grade on test

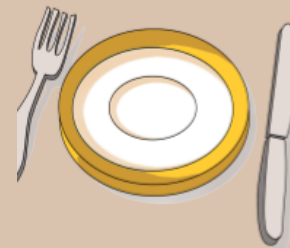
Amount of rainfall/Yield of corn per bushel



Sep 22-7:26 AM

Association

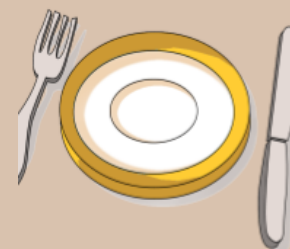
- The main purpose of data analysis with two variables is to investigate whether there is an association between them and to describe that association
- An association exists between two variables if a particular value for one variable is more likely to occur with certain values of the other variable



Sep 22-7:26 AM

For instance...

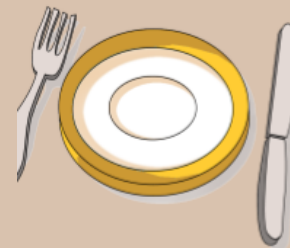
- Having a college GPA above 3.5 is more likely with a high school GPA of 4.0 than a high school GPA of 3.0.
- So high school GPA and college GPA have an association.



Sep 22-7:26 AM

Where we are headed...

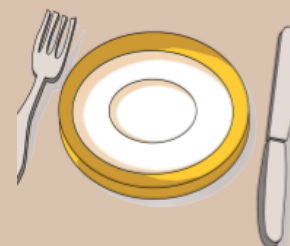
- 3.1 will be looking at association between categorical variables
- 3.2 and 3.3 will be looking at association between quantitative variables



Sep 22-7:26 AM

Contingency Tables

- Useful for looking at two categorical variables
- Think back to the GSS – how happy someone is was associated with his/her marital status



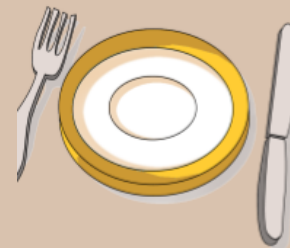
Sep 22-7:26 AM

Contingency Tables

- Displays two categorical variables
- The rows list the categories of one variable
- The columns list the categories of the other variable
- Entries in the table are frequencies

Food Type	Pesticides	
	Yes	No
Organic	29	98
Conventional	19485	7086

cell ↑



Sep 22-7:26 AM

Contingency Tables

Happiness /Income	Not Too Happy	Pretty Happy	Very Happy	Total
Above Average		294	272	615
Average	131		454	1420
Below Average		527	185	920
Total	388	1656		2955

Sep 22-7:26 AM

Contingency Tables

- What are the explanatory and response variables?

Happiness /Income	Not Too Happy	Pretty Happy	Very Happy	Total
Above Average	49	294	272	615
Average	131	835	454	1420
Below Average	208	527	185	920
Total	388	1656	911	2955

Sep 22-7:26 AM

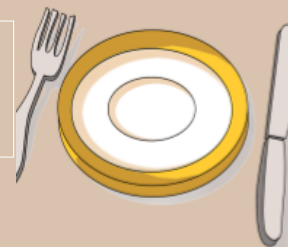
Contingency Tables

TABLE 3.1: Frequencies for Food Type and Pesticide Status.

The row totals and the column totals are the frequencies for the categories of each variable. The counts inside the table give information about the association.

Food Type	Pesticide Status		Total
	Present	Not Present	
Organic	29	98	127
Conventional	19485	7086	26571
Total	19514	7184	26698

What is the explanatory variable?
What is the response variable?



Sep 22-7:26 AM

Contingency Tables

Calculate conditional and marginal proportions

Food Type	Pesticides		Total
	Yes	No	
Organic	29	98	127
Conven.	19485	7086	26571
Total	19514	7184	26698

TABLE 3.2: Conditional Proportions on Pesticide Status, for Two Food Types.

These treat pesticide status as the response variable. The sample size n in a row shows the total on which the conditional proportions in that row were based.

Food Type	Pesticide Status		Total	n
	Present	Not Present		
Organic	0.23	0.77	1.000	127
Conventional	0.73	0.27	1.000	26571

Sep 22-7:26 AM

Contingency Tables

Use your table!

What proportion of organic foods contain pesticides? (conditional)

0.23

What proportion of conventionally grown foods contain pesticides? (conditional)

0.73

What proportion of all sampled items contain pesticide residuals? (marginal)

0.731

Sep 22-7:26 AM

Brief Hiatus

- What are some real world applications of this contingency table stuff?
- Drug testing anyone?

Sep 22-7:26 AM

“Marion Jones Pleads Guilty in Drug Case, Retires”



Marion Jones speaks to the media Oct. 5, 2007, at federal court in White Plains, N.Y., after pleading guilty to lying to a federal agent about her steroid use. Getty Images All Things Considered, October 5, 2007 by Michele Norris

<http://www.npr.org/templates/story/story.php?storyId=15060426>

Sep 22-7:26 AM

“IOC will consider leaving Marion Jones' Sydney 100-meter gold medal spot vacant”

By Stephen Wilson, AP Sports Writer, Nov 15, 2007

http://www.usatoday.com/sports/olympics/2007-11-15-966036254_x.htm

MADRID, Spain — Future record books may show there was no winner of the women's 100 meters at the 2000 Sydney Olympics.

IOC officials said Thursday they are considering the unprecedented step of leaving the gold-medal spot vacant following Marion Jones' confession that she used performance-enhancing drugs.

"That could be a solution," said IOC board member Denis Oswald, part of the three-man disciplinary commission dealing with the Jones case.

Although the original second-place finisher normally would be upgraded to the gold, the International Olympic Committee is reluctant to give the medal to Greek sprinter Katerina Thanou because she was caught up in a doping scandal at the 2004 Athens Games.

Sep 22-7:26 AM

“Texas hires firm to test school athletes for steroids: Drug Free Sport gets \$5.6 million, two-year deal to initiate program in the next few weeks”

By GARY SCHARRER Jan. 23, 2008, 12:07AM Houston Chronicle Austin Bureau

<http://www.chron.com/disp/story.mpl/headline/metro/5476779.html>

Testing will start within weeks and will affect between 40,000 and 50,000 student athletes by the end of the 2009 school year, said Kim Rogers, spokeswoman for the University Interscholastic League, which governs Texas public school sports.

Texas lawmakers last year mandated the random tests for steroid use, and it will be the largest such program in the country.....

The random testing program will cover about 400 of the state's nearly 1,300 high schools and about 3 percent per year of the more than 700,000 students who participate in sports.

The testing is expected to start in February.

Sep 22-7:26 AM

“Drug Testing Set For Student Athletes In Illinois”

Reporting Dorothy Tucker CHICAGO (CBS) Jan 15, 2008 6:23 pm US/Central

As CBS 2's Dorothy Tucker explains, now some high school athletes in Illinois will have to take another one, a drug test.

This year's state finals will be the last games students play without being tested for steroids.

Next year, whether it's basketball, football, or any other sport, whether it's girls or boys, every athlete competing in sectionals, regionals or championship games will be subject to random testing of banned drugs.

Sep 22-7:26 AM

Do Drug Tests Work? What can go wrong?

- **Do they change student behavior? Not clear!** Discussion of how Indiana students get around frequent drug testing in “Why Drug Tests Flunk” at http://dir.salon.com/story/mwt/feature/2002/04/22/drug_testing/ (put salt in urine sample, or a strand of hair coated with hair spray)
- **What about students who get a positive test result? Is it certain they have used drugs? What about students who get a negative test result? Are we sure they are clean? Is it possible the test is wrong? Is it likely?**

Sep 22-7:26 AM

We'll look at the second issue:

- How likely is it that a student is falsely accused?
- How likely is it that a cheating student is missed?
- Mathematically, what are these questions about?
(*Answer: Proportions, percentages, or probability.*)

Sep 22-7:26 AM

What Do the Answers Depend On?

- Accuracy of the test
- Frequency of drug use
- Accuracy: Two different components
- *Sensitivity* is proportion of times a user shows up positive
- *Specificity* is proportion of times a nonuser shows up negative
- Why are sensitivity and specificity not the same? Can you make both 100%?

Sep 22-7:26 AM

Data:

- **Sensitivity:** We'll start with 90%
- **Specificity:** We'll start with 95%
- **Prevalence of Drug Use:** We'll start with 1%

Questions:

- **What proportion of students who test positive really are using drugs?**
- **What proportion of students who test negative really are clean?**

Sep 22-7:26 AM

Data:

- **Sensitivity:** We'll start with 90%
- **Specificity:** We'll start with 95%
- **Prevalence of Drug Use:** We'll start with 1%

Questions:

- **What proportion of students who test positive really are using drugs?**
- **What proportion of students who test negative really are clean?**

Sep 22-7:26 AM

Another Example: Mammograms

- In the 1990s, the American Medical Association and the National Academy of Sciences debated whether women in their 40 should have mammograms.
- What were the issues?
- Whether the danger from the mammogram (and possible panic or surgery after a positive result) was worth the risk
- How accurate they were

Sep 22-7:26 AM

Mammogram: Question and Data

The Question:

- If a woman has a positive mammogram, what is the chance she has breast cancer?

The Data:

- Sensitivity about 75%, specificity about 90%
- Prevalence 0.23% for women in their 40's

The Result:

- There's less than a 2% chance that a positive mammogram means cancer.

Sep 22-7:26 AM

What About Mammograms for Older Women?

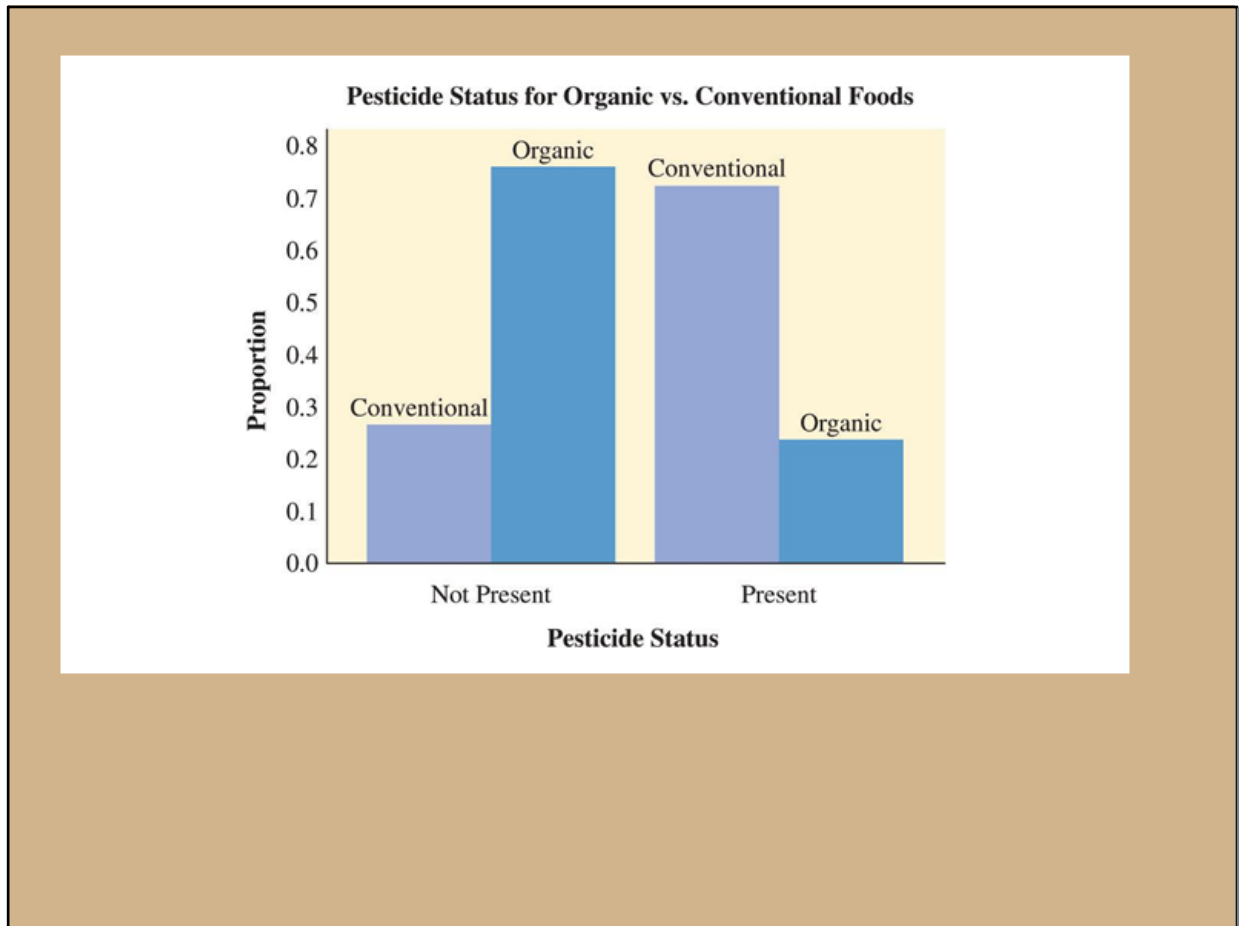
- What difference does it make that the chance of cancer increases with age?
- At some prevalence level, the benefits of the mammogram outweigh the risks
- There has been no debate about the value of mammograms for women in their 50s and above

Sep 22-7:26 AM

Last Bit...

- Use side by side bar charts to show conditional proportions
- Allows for easy comparison of the explanatory variable with respect to the response variable

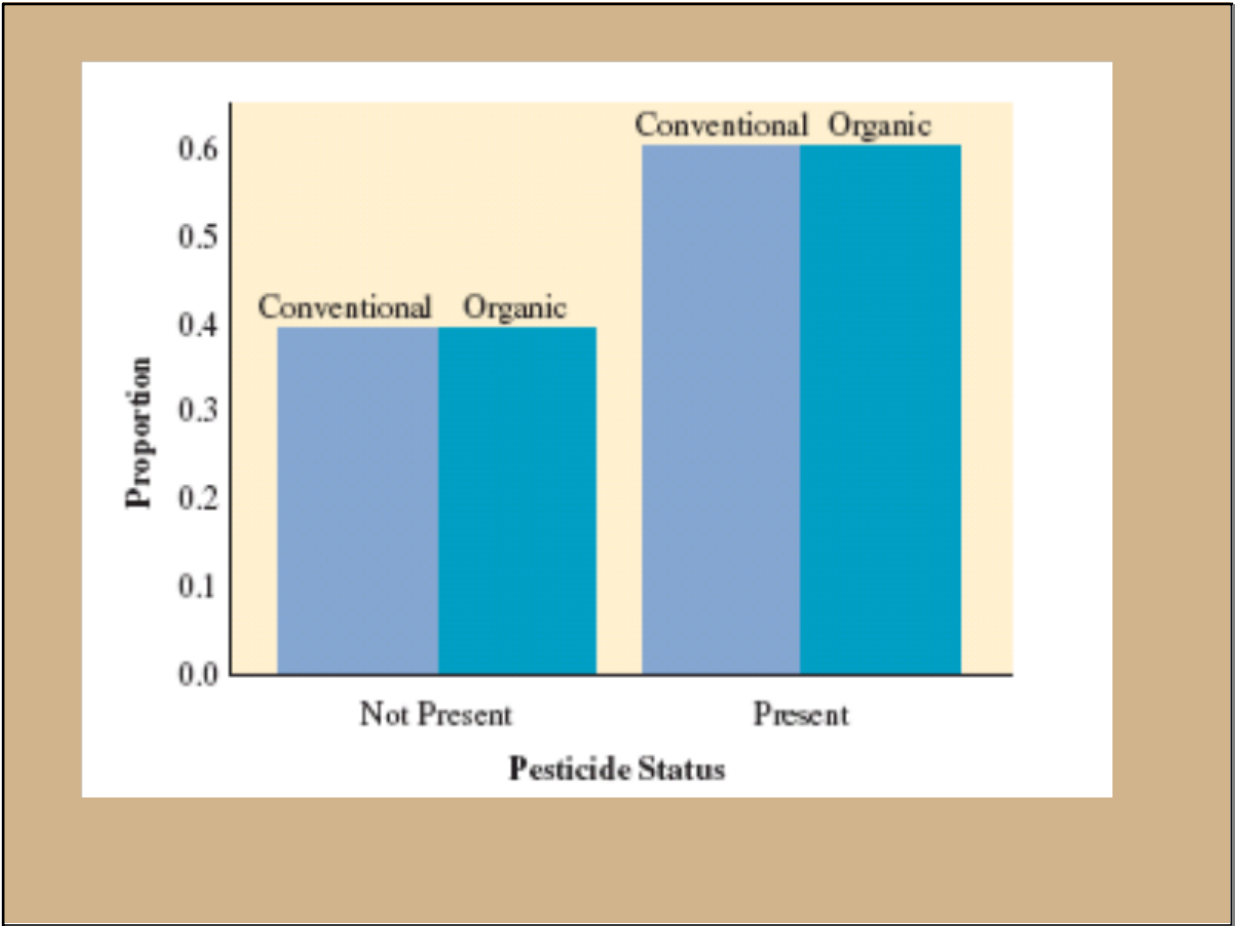
Sep 22-7:26 AM



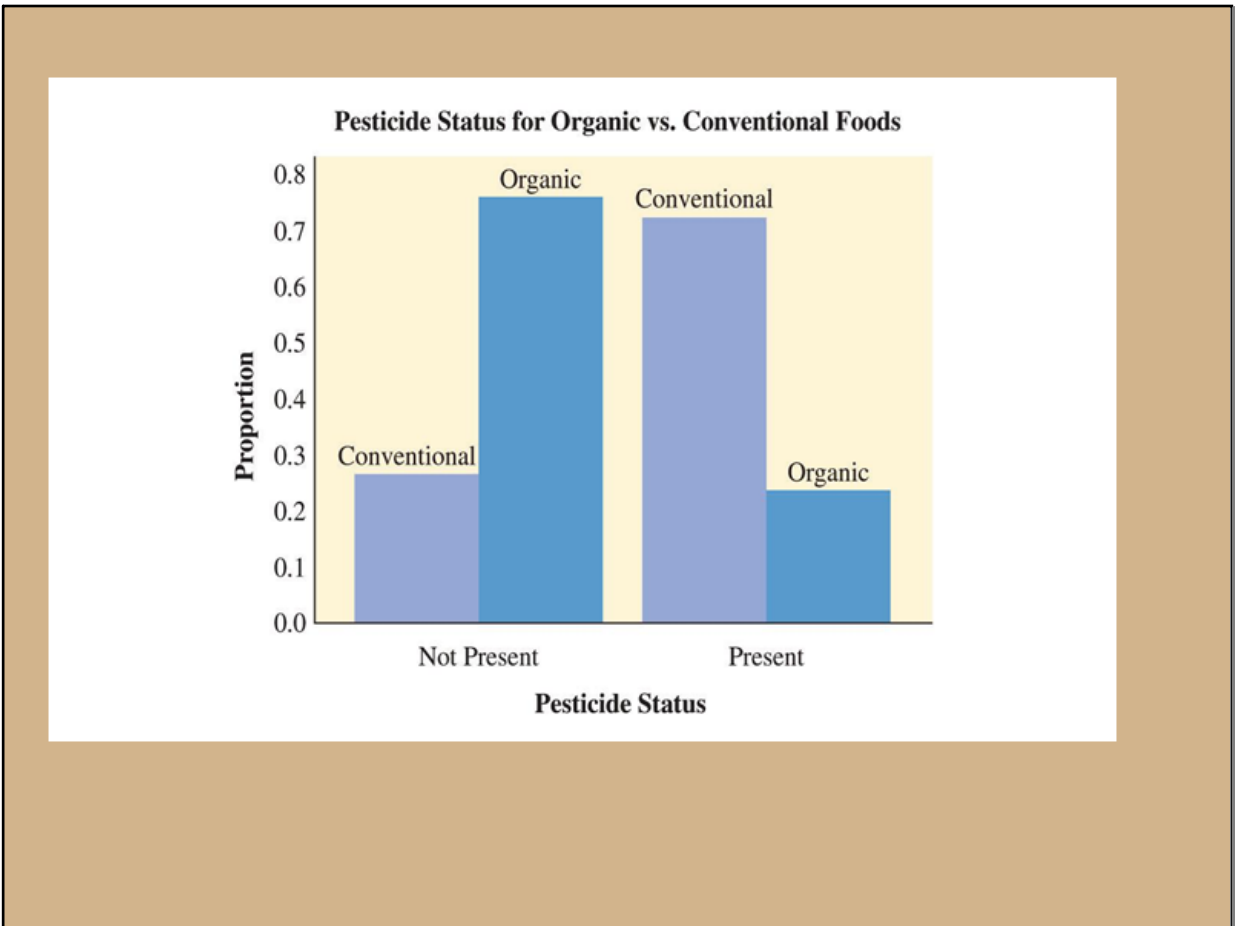
Sep 22-7:26 AM

-If there was no association between organic and conventional foods, then the proportions for the response variable categories would be the same for each food type

Sep 22-7:26 AM



Sep 22-7:26 AM



Sep 22-7:26 AM

Homework

- Page 100
- #1, 2, 3, 5, 6, 7, and 9

Sep 22-7:26 AM

Sep 22-7:41 AM