Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Coverage, adherer bias, and clofibrate example.
2. More about confounding factors.
3. Lefties example.
4. One-sample formulas for numerical and quantitative data.
5. Comparing two proportions using numerical and visual summaries, good or bad year example.
6. Comparing 2 proportions with CIs + testing using simulation, dolphin example.

http://www.stat.ucla.edu/~frederic/13/W24 .
HW2 due Mon, Feb12, 1159pm. 2.3.15, 3.3.18, and 4.1.23.
Finish chapter 4.
Midterm is Mon Feb26 in class. Bring a pencil or pen, and a calculator.
On the exam, you cannot use computers or ipads or phones or anything that can surf the web or do email.

1. Coverage, adherer bias and Clofibrate example.

Surveys are observational.

- Coverage is a common issue. Coverage is the extent to which the people you sampled from represent the overall population. A survey at a fancy research hospital in a wealthy neighborhood may yield patients with higher incomes, higher education, etc.
- Non-response bias is another common problem. Poor coverage means the people getting the survey do not represent the general population. Non-response bias means that out of the people you gave the survey to, the people actually filling it out and submitting it are different from the people who did not.
- Same exact issues in web surveys.

Coverage, adherer bias, and Clofibrate example.

Non-response bias is similar to adherer bias, in experiments.
A drug called clofibrate was tested on 3,892 middle-aged men with heart trouble. It was supposed to prevent heart attacks.

1,103 assigned at random to take clofibrate,

2,789 to placebo (lactose) group.

Subjects were followed for 5 years.

Is this an experiment or an observational study?

Clofibrate	patients who died during followup
adherers	15%
non-adherers	25%
total	20%

Coverage, adherer bias, and Clofibrate example.

Non-response bias is similar to adherer bias, in experiments.
A drug called clofibrate was tested on 3,892 middle-aged men with heart trouble. It was supposed to prevent heart attacks.

1,103 assigned at random to take clofibrate,

2,789 to placebo (lactose) group.

Subjects were followed for 5 years.

Is this an experiment or an observational study?

It is an experiment. Does Clofibrate work?

Clofibrate	patients who died during followup
adherers	**15%**
non-adherers	**25%**
total	20%

Clofibrate	patients who died during followup
adherers	**15%**
non-adherers	**25%**
total	20%

Placebo	
adherers	15%
nonadherers	28%
total	21%

Those who took clofibrate did much better than those who didn't keep taking clofibrate. Does this mean clofibrate works?

Clofibrate	patients who died during followup
adherers	15%
non-adherers	25%
total	20%

Placebo	
adherers	**15%**
nonadherers	**28%**
total	21%

Those who adhered to placebo also did much better than those who stopped adhering.

Clofibrate	patients who died during followup
adherers	15%
non-adherers	25%
total	**20%**

Placebo	
adherers	15%
nonadherers	28%
total	**21%**

All in all there was little difference between the two groups.

Clofibrate	patients who died during followup
adherers	15%
non-adherers	25%
total	**20%**

Placebo	
adherers	15%
nonadherers	28%
total	**21%**

Adherers did better than non-adherers, not because of clofibrate, but because they were healthier in general. Why?

Clofibrate	patients who died during followup
adherers	15%
non-adherers	25%
total	**20%**

Placebo	
adherers	15%
nonadherers	28%
total	**21%**

Adherers did better than non-adherers, not because of clofibrate, but because they were healthier in general. Why?

- adherers are the type to engage in healthier behavior.
- sick patients are less likely to adhere.

2. More about confounding factors.

- By a confounding factor, we mean an alternative explanation that could explain the apparent relationship between the two variables, even if they are not causally related. Typically this is done by finding another difference between the treatment and control group. For instance, different studies have examined smokers and non-smokers and have found that smokers have higher rates of liver cancer. One explanation would be that smoking causes liver cancer. But is there any other, alternative explanation?
- One alternative would be that the smokers tend to drink more alcohol, and it is the alcohol, not the smoking, that causes liver cancer.

More about confounding factors.

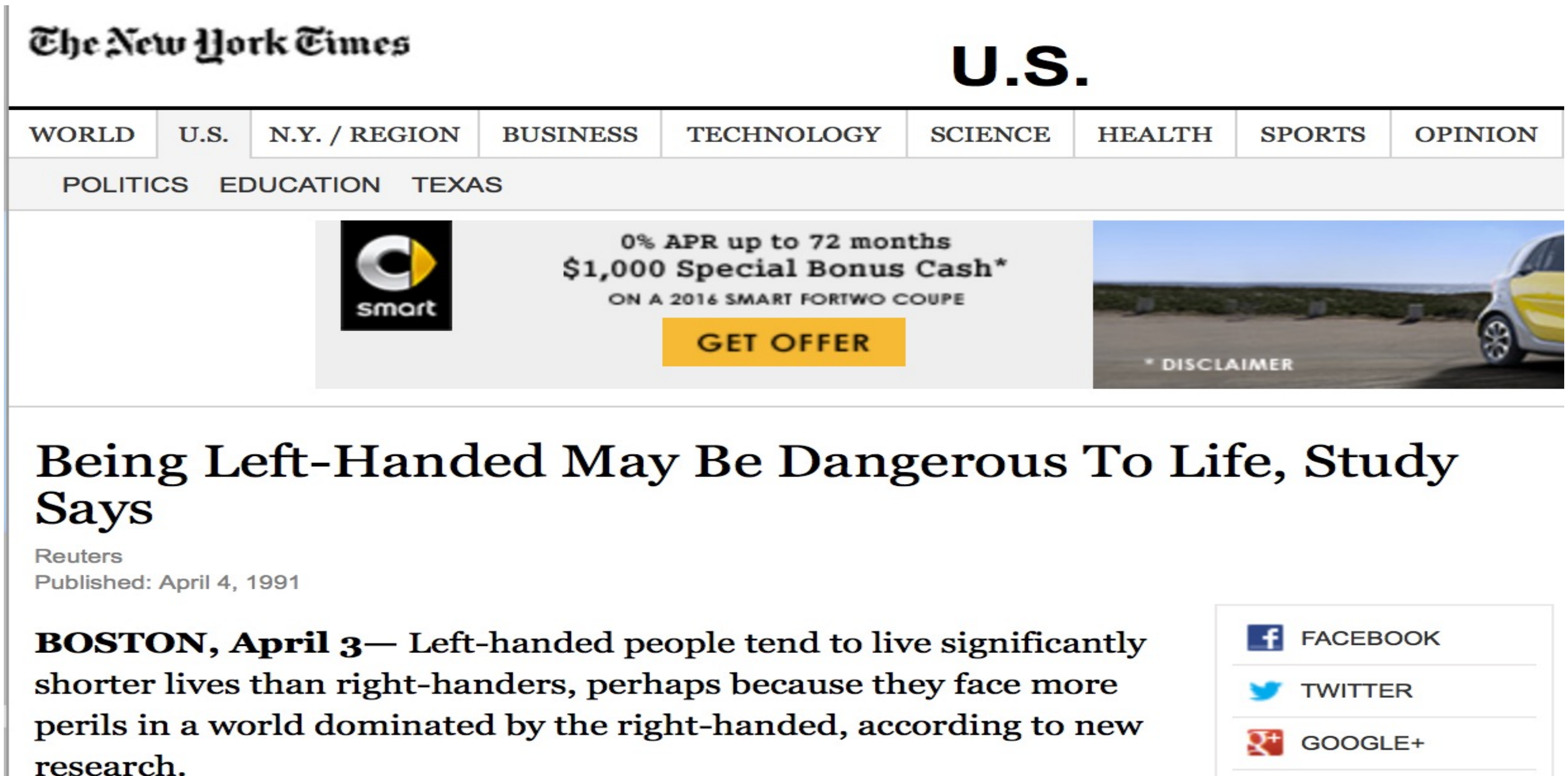
- Another plausible explanation is that the smokers are probably older on average than the non-smokers, and older people are more at risk for all sorts of cancer than younger people.
- Another might be that smokers engage in other unhealthy activities more than non-smokers.
- Note that if one said that “smoking makes you want to drink alcohol which causes liver cancer,” that would not be a valid confounding factor, since in that explanation, smoking effective is causally related to liver cancer risk.

3. Lefties example.

- A confounding factor must be plausibly linked to both the explanatory and response variables. So for instance saying “perhaps a higher proportion of the smokers are men” would not be a very convincing confounding factor, unless you have some reason to think gender is strongly linked to liver cancer.
- Another example: left-handedness and age at death. Psychologists Diane Halpern and Stanley Coren looked at 1,000 death records of those who died in Southern California in the late 1980s and early 1990s and contacted relatives to see if the deceased were righthanded or lefthanded. They found that the average ages at death of the lefthanded was 66, and for the righthanded it was 75. Their results were published in prestigious scientific journals, Nature and the New England Journal of Medicine.

Lefties example.

All sorts of causal conclusions were made about how this shows that the stress of being lefthanded in our righthanded world leads to premature death.



The New York Times

U.S.

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

POLITICS EDUCATION TEXAS

Being Left-Handed May Be Dangerous To Life, Study Says

Reuters
Published: April 4, 1991

BOSTON, April 3— Left-handed people tend to live significantly shorter lives than right-handers, perhaps because they face more perils in a world dominated by the right-handed, according to new research.

FACEBOOK
TWITTER
GOOGLE+

Lefties example.

- Is this an observational study or an experiment?

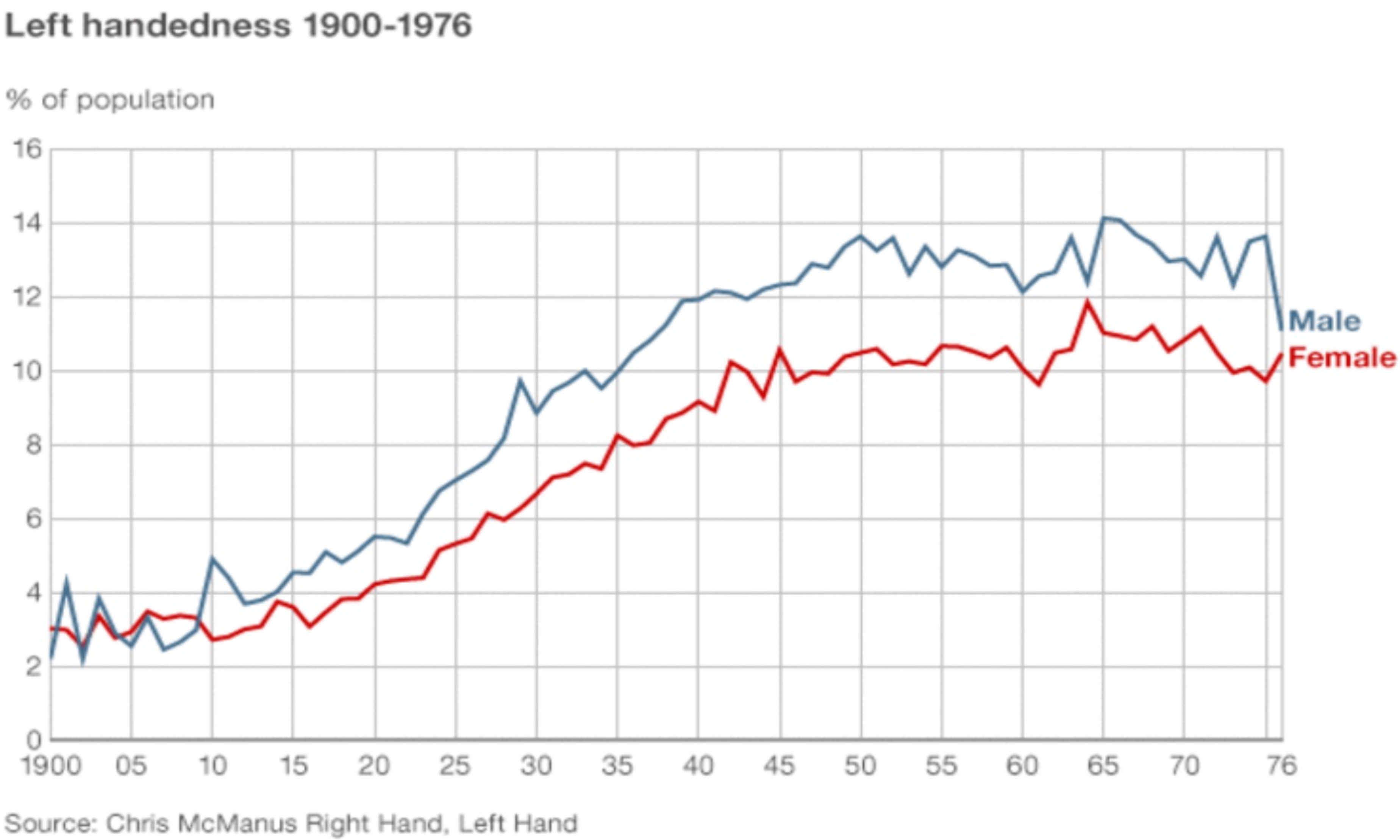
Lefties example.

- Is this an observational study or an experiment?

It is an observational study.

- Are there plausible confounding factors you can think of?

Lefties example.



4. Formulas for CIs for one variable, quantitative or categorical.

if the observations are iid and n is large, then

P(μ is in the range $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$) ~ 95%.

and since s ~ σ when n is large, 95% CI is

$\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

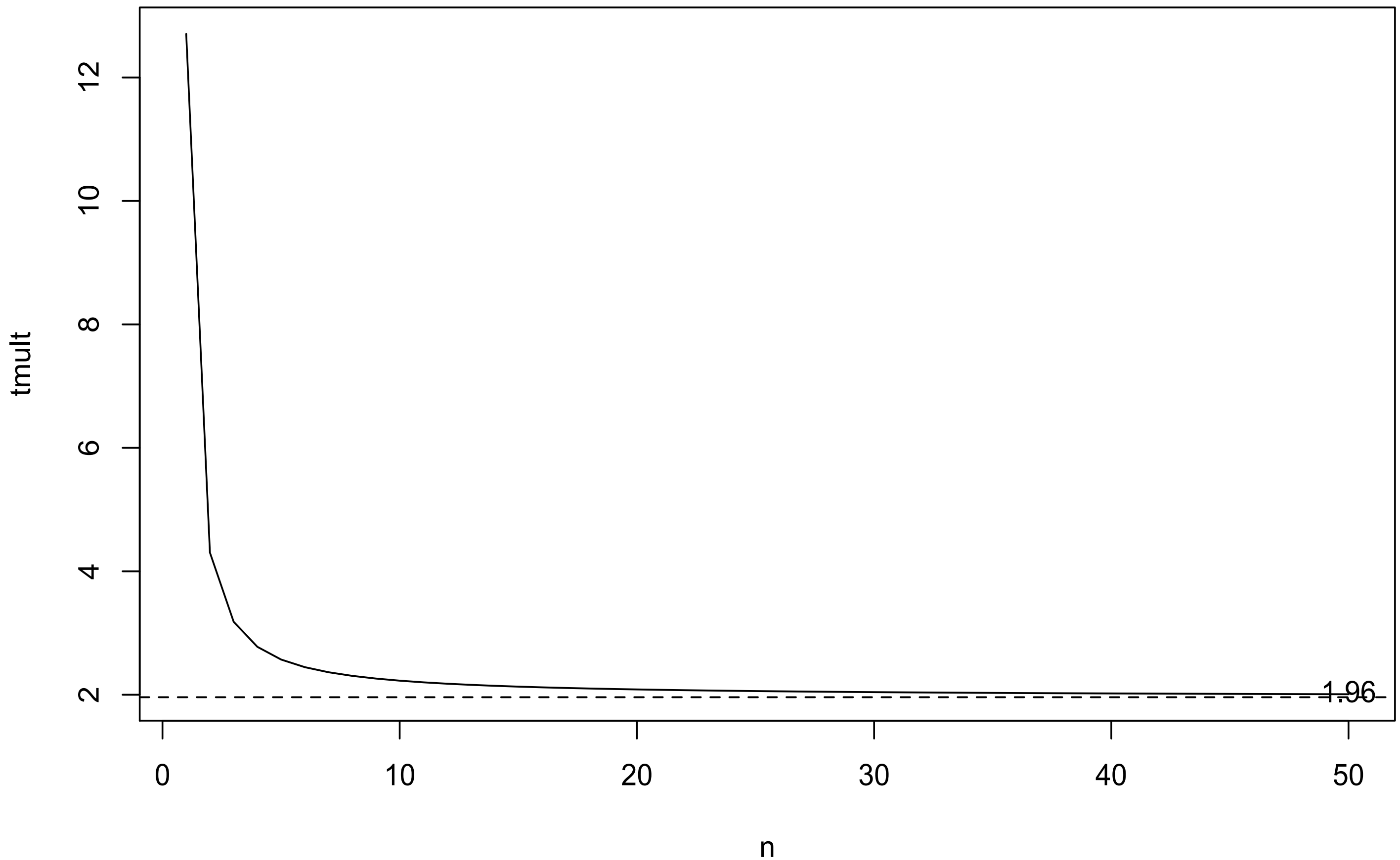
If the obs. are iid and normal and σ is unknown, then

even if n is small,

P(μ is in the range $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$) ~ 95%.

where t_{mult} depends on n.

$\bar{x}$ +/- t_{mult} $s/\sqrt{n}$.



t_{mult} gets really close to 1.96 when n gets larger than about 30, so for this class we will use the rule of thumb

$n \geq 30$ is large, for quantitative data. For categorical, at least 10 of each type in your sample will be the rule of thumb.

4. Formulas for CIs for one variable, quantitative or categorical.

Note that for quantitative variables, in the 95% CI formula

$$\bar{x} \text{ +/- } \mathbf{1.96\ s/\sqrt{n}},$$

The quantity $s / \sqrt{n}$ is called the SE for the mean.

For categorical data, the population is never normal!
View the values as 0 or 1. Then

$\hat{p} = \bar{x}$, and $s = \sqrt{[\hat{p}(1-\hat{p})]}$. So the formula for a 95% CI is

$$\hat{p} \text{ +/- } 1.96 \sqrt{[\hat{p}(1-\hat{p})/n]}.$$

Here large n means ≥ 10 of each type in the sample.

Unit 2. Comparing Two Groups

- In Unit 1, we learned the basic process of statistical inference using tests and confidence intervals. We did all this by focusing on a single proportion.
- In Unit 2, we will take these ideas and extend them to comparing two groups. We will compare two proportions, two independent means, and paired data.

5. Comparing two proportions using numerical and visual summaries, and the good or bad year example.

Section 5.1

Example 5.1: Positive and Negative Perceptions

- Consider these two questions:
  - Are you having a good year?
  - Are you having a bad year?

- Do people answer each question in such a way that would indicate the same answer? (e.g. Yes for the first one and No for the second.)

Positive and Negative Perceptions

- Researchers questioned 30 students (randomly giving them one of the two questions).
- They then recorded if a positive or negative response was given.
- They wanted to see if the wording of the question influenced the answers.

Positive and negative perceptions

- Observational units
  - The 30 students
- Variables
  - Question wording (good year or bad year)
  - Perception of their year (positive or negative)
- Which is the explanatory variable and which is the response variable?
- Is this an observational study or experiment?

Raw Data in a Spreadsheet

Individual	Type of Question	Response
1	Good Year	Positive
2	Good Year	Negative
3	Bad Year	Positive
4	Good Year	Positive
5	Good Year	Negative
6	Bad Year	Positive
7	Good Year	Positive
8	Good Year	Positive
9	Good Year	Positive
10	Bad Year	Negative
11	Good Year	Negative
12	Bad Year	Negative
13	Good Year	Positive
14	Bad Year	Negative
15	Good Year	Positive

Individual	Type of Question	Response
16	Good Year	Positive
17	Bad Year	Positive
18	Good Year	Positive
19	Good Year	Positive
20	Good Year	Positive
21	Bad Year	Negative
22	Good Year	Positive
23	Bad Year	Negative
24	Good Year	Positive
25	Bad Year	Negative
26	Good Year	Positive
27	Bad Year	Negative
28	Good Year	Positive
29	Bad Year	Positive
30	Bad Year	Negative

Two-Way Tables

- A **two-way table** organizes data
  - Summarizes *two* categorical variables
  - Also called contingency table
- Are students more likely to give a positive response if they were given the good year question?

	Good Year	Bad Year	Total
Positive response	15	4	19
Negative response	3	8	11
Total	18	12	30

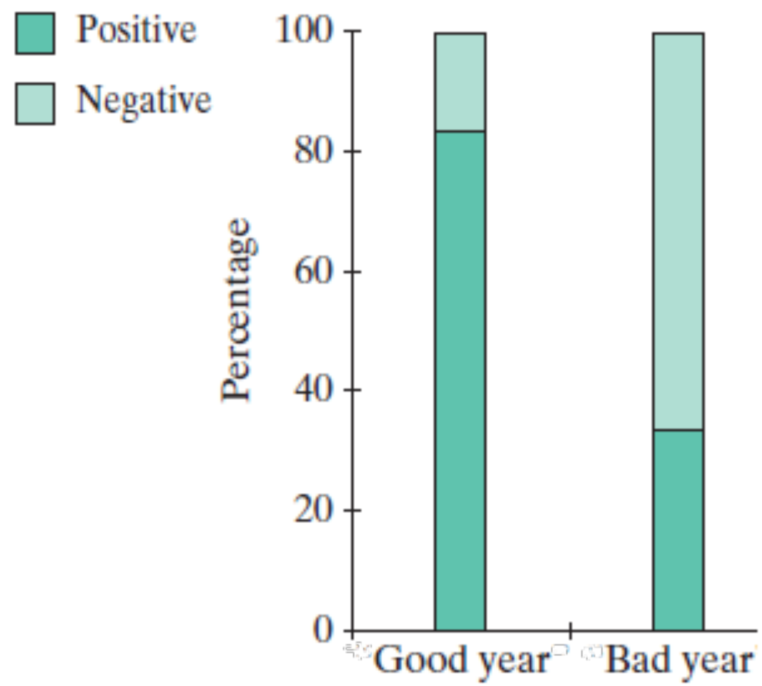
Two-Way Tables

- Conditional proportions will help us better determine if there is an association between the question asked and the type of response.
- We can see that the subjects with the positive question were ***more likely*** to respond positively.

	Good Year	Bad Year	Total
Positive response	15/18 ≈ 0.83	4/12 ≈ 0.33	19
Negative response	3	8	11
Total	18	12	30

Segmented Bar Graphs

- We can also use segmented bar graphs to see this association between the "good year" question and a positive response.



Statistic

- The statistic we will mainly use to summarize this table is the difference in proportions of positive responses is 0.83 – 0.33 = 0.50.

	Good Year	Bad Year	Total
Positive response	15 (83%)	4 (33%)	19
Negative response	3	8	11
Total	18	12	30

Another Statistic

- Another statistic that is often used, called **relative risk**, is the ratio of the proportions: 0.83/ 0.33 = 2.5.
- We can say that those who were given the good year question were 2.5 times as likely to give a positive response.

	Good Year	Bad Year	Total
Positive response	15 (83%)	4 (33%)	19
Negative response	3	8	11
Total	18	12	30

6. Comparing two proportions with CIs and testing using simulation, dolphin example.

Section 5.2

Swimming with Dolphins

Example 5.2

Swimming with Dolphins

Is swimming with dolphins therapeutic for patients suffering from clinical depression?

- Researchers Antonioli and Reveley (2005), in British Medical Journal, recruited 30 subjects aged 18-65 with a clinical diagnosis of mild to moderate depression
- Discontinued antidepressants and psychotherapy 4 weeks prior to and throughout the experiment
- 30 subjects went to an island near Honduras where they were randomly assigned to two treatment groups

Swimming with Dolphins

- Both groups engaged in one hour of swimming and snorkeling each day
- One group swam in the presence of dolphins and the other group did not
- Participants in both groups had identical conditions except for the dolphins
- After two weeks, each subjects' level of depression was evaluated, as it had been at the beginning of the study
- The response variable is whether or not the subject achieved substantial reduction in depression

Swimming with Dolphins

Null hypothesis: Dolphins do not help.

- Swimming with dolphins is not associated with substantial improvement in depression

Alternative hypothesis: Dolphins help.

- Swimming with dolphins **increases** the probability of substantial improvement in depression symptoms

Swimming with Dolphins

- The parameter is the (long-run) difference between the probability of improving when receiving dolphin therapy and the prob. of improving with the control ($\pi_{\text{dolphins}} - \pi_{\text{control}}$)
- So we can write our hypotheses as:

H_0: $\pi_{\text{dolphins}} - \pi_{\text{control}} = 0.$

H_a: $\pi_{\text{dolphins}} - \pi_{\text{control}} > 0.$

or

H_0: $\pi_{\text{dolphins}} = \pi_{\text{control}}$

H_a: $\pi_{\text{dolphins}} > \pi_{\text{control}}$

(Note: we are not saying our parameters equal any certain number.)

Swimming with Dolphins

Results:

	Dolphin group	**Control group**	**Total**
Improved	10 (66.7%)	3 (20%)	13
Did Not Improve	5	12	17
Total	15	15	30

The difference in proportions of improvers is:

$\hat{p}_d - \hat{p}_c$ **= 0.667 – 0.20 = 0.467.**

Swimming with Dolphins

- There are two possible explanations for an observed difference of 0.467.
  - A tendency to be more likely to improve with dolphins (alternative hypothesis)
  - The 13 subjects were going to show improvement with or without dolphins and random chance assigned more improvers to the dolphins (null hypothesis)