


# Introduction to Statistics I

Instructor: Jodin Morey    moreyj@lemoyne.edu

## Previous Lecture

- ◆ Errors T1E/T2E
- ◆ How to minimize errors
- ◆ Power 



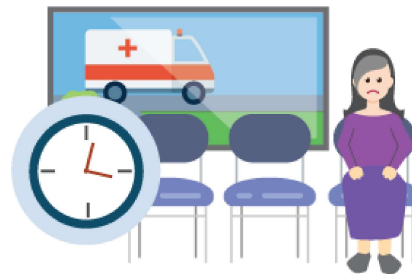
## Topic 6: Two-Way Tables

### Multiple Variables

#### Examples RQs:

- ◆ Do uninsured patients wait longer at emergency rooms than the insured?

Two Variables: insurance status & wait time.

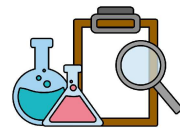


- ◆ Is average waiting time at a hospital more than 30 mins?

One Variable: waiting time.

- ◆ Do doctors order more tests for women than for men?

Two Vars: Avg # tests ordered for women and for men.



- ◆ Is average bill more than \$5,000 for an emergency room visit?

One Variable: amount of bill.



#### Recall: Explanatory & Response Vars

From First Example: Explanatory - Insurance Status

Response - Wait Times

# Catagorical Vars

Let's explore the relationships between *two* catagorical vars.

**Example RQ:** Are women more likely to report feeling rushed than men? Explanatory/Response?

Explanatory Var: male or female.  
Response Var: how frequently they feel rushed.



## Two-Way Tables

Two-way tables are used to examine relationship between **two categorical vars**.  
When gathering data, we classify obs-units **two ways**.

**Gender? How frequently do you feel rushed?**

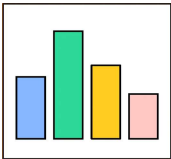
	Males	Females	Total
Always	116	188	304
Sometimes	239	284	513
Never	82	78	160
Total	427	550	977

Explanatory categories listed across **top/columns**.  
Response categories listed down **side/rows**.

# in each cell is # of obs-units who fit into both row and column categories.

## Marginal Distr's

Recall: The distr of a categorical var is the categories of that var, and the proportions in each category.



**Marginal distr** of a categorical var in a two-way table is the distr of that var, ignoring the other.  
In particular, it's the distr (proportions) of the values in the "margins" (i.e., the totals).

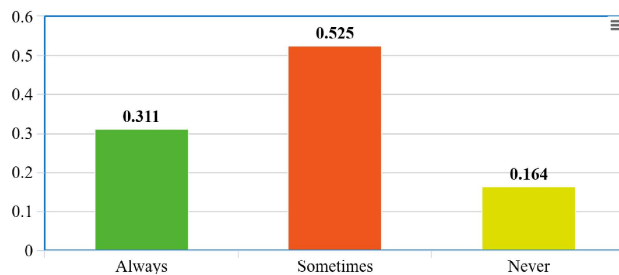
**"How frequently do you feel rushed" marginal distr** (ignoring male/female):

	Males	Females	Total
Always	116	188	<b>304</b>
Sometimes	239	284	<b>513</b>
Never	82	78	<b>160</b>
Total	427	550	<b>977</b>

	Total
Always	<b>304</b>
Sometimes	<b>513</b>
Never	<b>160</b>
<b>Total</b>	<b>977</b>

### Proportions:

Always:	$\frac{304}{977} \approx 0.3112$
Sometimes:	$\frac{513}{977} \approx 0.5251$
Never:	$\frac{160}{977} \approx 0.1638$



! Proportions for a distr always sum to 1.

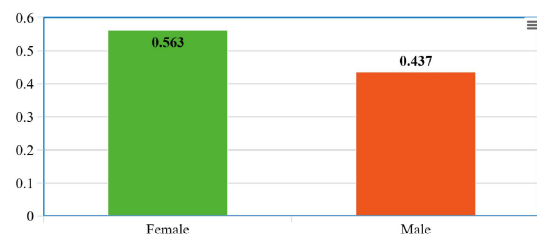
**Gender marginal distr** (ignoring frequency):

	Males	Females	Total
Always	116	188	304
Sometimes	239	284	513
Never	82	78	160
<b>Total</b>	<b>427</b>	<b>550</b>	<b>977</b>

	Males	Females	Total
<b>Total</b>	<b>427</b>	<b>550</b>	<b>977</b>

### Proportions:

Males:	$\frac{427}{977} = 0.4371$
Females:	$\frac{550}{977} = 0.5630$



## Conditional Distr's

The **conditional distribution** of a categorical var in a two way table is the distr of that var, restricted to *just one category*.

**Example:** What is conditional distr of "how frequently do you feel rushed" when restricted to just females?

We're giving the distr of one var (frequency) *conditioning* on a category (female).

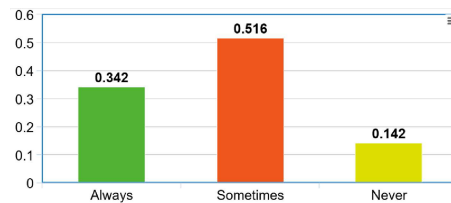
### Feeling Rushed Conditional Distribution

	Males	Females	Total
Always	116	<b>188</b>	304
Sometimes	239	<b>284</b>	513
Never	82	<b>78</b>	160
<b>Total</b>	<b>427</b>	<b>550</b>	<b>977</b>

	Females
Always	<b>188</b>
Sometimes	<b>284</b>
Never	<b>78</b>
<b>Total</b>	<b>550</b>

### Females

Always:	$\frac{188}{550} \approx 0.3418$
Sometimes:	$\frac{284}{550} \approx 0.5164$
Never:	$\frac{78}{550} \approx 0.1418$



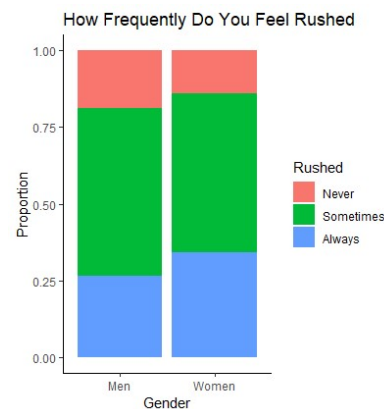
## Both Conditional Distributions of Feeling Rushed

	Males	Females
Always:	$\frac{116}{427} \approx 0.2717$	$\frac{188}{550} \approx 0.3418$
Sometimes:	$\frac{239}{427} \approx 0.5597$	$\frac{284}{550} \approx 0.5164$
Never:	$\frac{82}{427} \approx 0.1920$	$\frac{78}{550} \approx 0.1418$

## Segmented Bar Graph

Bar graphs are used to compare distr's of response var across categories of explanatory var.

We see a difference in conditional distr's, where women are more likely to be "always" rushed than males.



## Relationship Between Vars

### Independence

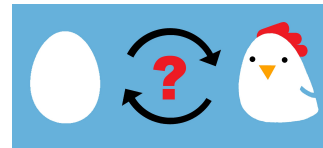
If the response var distr is (approx) the same across all categories of the explanatory var, then the explanatory var **has no effect** on the distr of the response var.

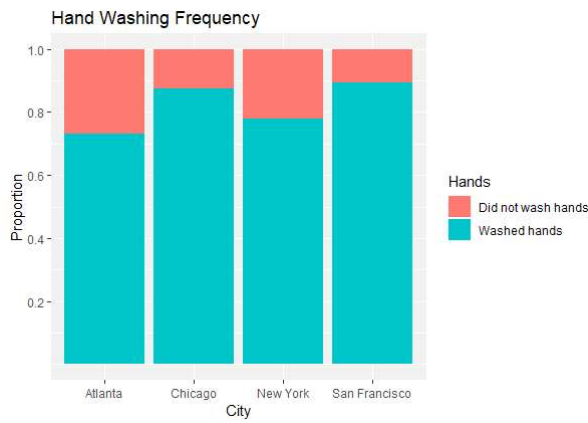
In this case, the vars are **independent**.

What if the distr of the response var is different across some categories of the explanatory var?

In this case, the vars have a **relationship**. However, this doesn't necessarily mean the vars have a cause-and-effect relationship (for this, we'd need an experiment).

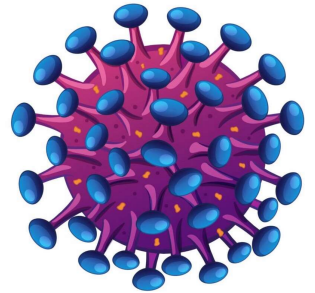
Do **city** and **hand washing frequency** have a relationship?





## Another Example: AZT & HIV

An early study, aimed at preventing maternal transmission of AIDS to infants, gave the drug AZT to pregnant, HIV infected patients. Roughly half of them were randomly assigned to receive AZT, the others received a placebo. Of 180 babies whose mothers got AZT, 13 were infected. Of 183 babies in placebo group, 40 got HIV. Explanatory/Response?



**Explanatory:** Whether mother received AZT or placebo

**Response:** Whether baby contracted HIV or not.

(Two-Way Table?) ...

**AZT / HIV Two-Way Table**

	AZT	Placebo	Total
No HIV	167	143	310
HIV	13	40	53
Total	180	183	363

## Conditional Dists

Distr of HIV for AZT babies?

$$\text{No HIV: } \frac{167}{180} \approx 0.9278$$

$$\text{HIV: } \frac{13}{180} \approx 0.07222$$

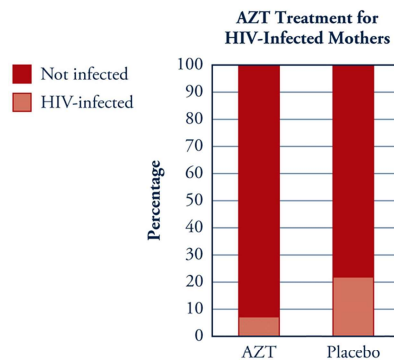
Distr of HIV for Placebo Babies?

$$\text{No HIV: } \frac{143}{183} \approx 0.7814$$

$$\text{HIV: } \frac{40}{183} \approx 0.2186$$

	AZT	Placebo
No HIV	0.9278	0.7814
HIV	0.07222	0.2186
Total	1.000	1.000

Segmented Bar Graph?



Does a baby's HIV status have a relationship w/which treatment the mother received?

## Relative Risk

How much better is it for a mother to receive AZT than to receive placebo?

Consider the proportion with disease for the two differing treatments.

Ratio of these proportions is called **Relative Risk** of contracting the disease (the risk if the less effective treatment is used. Here, placebo).

**Example:** Risk of HIV for AZT mothers: 0.072.

Risk of Placebo mothers: 0.219.

We can calculate ratio of these risks to determine **how many times worse** it is to take placebo, than to take AZT.

Place the higher proportion in numerator for relative risk.

Relative Risk:  $\frac{0.2186}{0.07222} \approx 3.027$

Risk of HIV for babies whose mother received placebo is **3.027 times higher** than for babies whose mother received AZT.

## Activity: 6-4

### What did we learn?

- ♦ Multiple Vars: Explanatory & Response Vars
- ♦ Two-way tables: marginal distr, conditional distr
- ♦ Segmented bar graphs
- ♦ Var independence: cause-and-effect relationships
- ♦ Relative risk

