

Stat 201: Introduction to Statistics

Standard 9: Variable Association – Contingency Table

Association of Variables – Two Categorical Variables

- **Response Variable** – this is our dependent variable, the outcome variable on which comparisons are made
- **Explanatory Variable** – this is our independent variable, the groups to be compared with respect to values on the response variable
- **Think “we use the explanatory variable to EXPLAIN what’s going on with the response variable.”**

More Definitions

- 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

Response Variable	this is our dependent variable, the outcome variable on which comparisons are made
Explanatory Variable	this is our independent variable, the groups to be compared with respect to values on the response variable
Association	exists between two variables if a particular value for one variable is more likely to occur with certain values of the other variable

Examples

- Example 1:
 - Response: Age of death (quantitative)
 - Explanatory: Cigarettes smoked per day (quantitative)
- The idea here is that an experimental unit's smoking status gives us some of the information about how long they will live
 - [Actuaries](#) do this sort of thing – evidence has shown that smoking decreases your life expectancy.

Examples

- Example 2:
 - Response: Happiness Level (categorical)
 - Explanatory: Income Level (quantitative)
- The idea here is that an experimental unit's income level gives us some of the information about how happy they are
 - Does money buy happiness? Some studies say you are happier making more money up to a point – the [current](#) estimate is \$75k.

Examples

- Example 3:
 - Response: Binge Drinking Status (categorical)
 - Explanatory: Gender (categorical)
- The idea here is that an experimental unit's gender gives us some of the information about whether or not they binge drink
 - [Recent studies](#) suggest a gender convergence in drinking habits with female alcohol consumption approaching that of males, but males currently edged out females on binge drinking

More Definitions

- 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

More Definitions

- “Evidence has shown that smoking more decreases your life expectancy.”
 - Here we say that there is an **association** between smoking and life expectancy.
- “Some studies say you are happier making more money up to a point – the current estimate is \$75k.”
 - Here we say that there is an **association** between income and happiness.
- “Males currently edged out females on binge drinking”
 - Here we say that there is an **association** between gender and binge drinking.

More Definitions

- **Contingency table:** A display for two categorical variables.
 - Its rows list the categories of one variable and its columns list categories of the other variable.

Contingency Table - Example

- Two Categorical Variables
 - Would you keep or turn in a \$100 if you found it on the library floor?
 - Do you recycle?

	Keep It	Turn It In	Total
No	17	8	25
Yes	30	34	64
Total	47	42	89

Contingency Table - Example

Counts

	Keep It	Turn It In	Total
No	17	8	25
Yes	30	34	64
Total	47	42	89

Percent:

(Divide each box by the overall total)

	Keep It	Turn It In	Total
No	17/89	8/89	25/89
Yes	30/89	34/89	64/89
Total	47/89	42/89	89/89

	Keep It	Turn It In	Total
No	19.1%	8.989%	28.09%
Yes	33.71%	38.2%	71.91%
Total	52.81%	47.19%	100%

Contingency Table - Example

Counts

	Keep It	Turn It In	Total
No	17	8	25
Yes	30	34	64
Total	47	42	89

Conditional Percent

(Divide each interior box by the row total)

	Keep It	Turn It In	Total
No	$17/25 = .68$	$8/25 = .32$	$25/25 = 1$
Yes	$30/64 = .4688$	$34/64 = .5313$	$64/64 = 1$
Total	$47/89 = .5281$	$42/89 = .4719$	$89/89 = 1$

	Keep It	Turn It In	Total
No	68%	32%	100%
Yes	46.88%	53.13%	100%
Total	52.81%	47.19%	100%

Contingency Table - Example

Counts		Keep It	Turn It In	Total
	No	17	8	25
	Yes	30	34	64
	Total	47	42	89

Percent		Keep It	Turn It In	Total
	No	19.1%	8.989%	28.09%
	Yes	33.71%	38.2%	71.91%
	Total	52.81%	47.19%	100%

Conditional Percent		Keep It	Turn It In	Total
	No	68%	32%	100%
	Yes	46.88%	53.13%	100%
	Total	52.81%	47.19%	100%

Contingency Table Example

Recycle\Money	Keep It	Turn It In	Total
No	68%	32%	100%
Yes	46.88%	53.13%	100%
Total	52.81%	47.19%	100%

- **Explanatory Variable**(rows): Recycling Status
- **Response Variable**(columns): Keep or Return Money

Contingency Table Example

Recycle\Money	Keep It	Turn It In	Total
No	68%	32%	100%
Yes	46.88%	53.13%	100%
Total	52.81%	47.19%	100%

- Does there appear to be an association between recycling and turning in money found on the floor?
 - Yes – by looking at the conditional percent contingency table it appears that a larger percent of people that recycle turn it in compared to those that keep it
 - For those who recycle more than half would turn it in as where only 32% of those who do not recycle would

Comparing Two Categorical Variables

Contingency table	A display for <u>two categorical variables</u> . Its rows list the categories of the explanatory variable and its columns list categories of the response variable.
Contingency table (Count)	Provides the counts of each categorical cross section
Contingency table (Percent)	Provides the total percentage of each categorical cross section. We divide each box by the total.
Contingency table (Conditional Percent)	Provides the conditional percentage of each categorical cross section. We divide each box by the row total.

Contingency Table

- StatCrunch Commands

Stat → Tables → Contingency Table → w/ data

→ Select your explanatory variable as the row variable

→ Select your response variable as the column variable

→ Choose which display you'd like

(**count**, percent of total(**percent**), row

percent(**conditional**)) → Compute!