

Applied Epidemiology I: Summary statistics, tables and interpreting results

Enoch Yi-Tung Chen

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet

December 3, 2020

Acknowledgements

This course material is based on my learning from Anastasia Lam's teachings in last year's Applied Epidemiology I lab sessions, and readings from *Epidemiology* by Gordis [1], *A First Course in Probability and Statistics* by Goldsman and Goldsman [2], *Principles of Biostatistics* by Pagano and Gauvreau [3], and *Biostatistics I* by Gabriel and Frumento [4]. I especially want to thank Marlene Stratmann for reviewing the slides and Prof. Paul Dickman for providing me with suggestions to improving the teaching.

Outline

① Summary statistics

Measures of Central Tendency:

mean, median, mode

Measures of Dispersion: range,
IQR, variance, standard
deviation

② Tables

Bad example

Basics of making tables

One-way tables

Two by two tables

Stata tool for Epidemiology

③ Basic Epidemiology terms

Rate vs. proportion

Risk, risk difference, risk ratio

Odds, odds ratio

④ Interpreting results

Principles

Ratio $>$ or $<$ 1

More examples

⑤ Calculate ratios using Stata

Risk ratio

Odds ratio

Incidence rate ratio

⑥ References

Summary statistics:

Measures of Central Tendency: mean, median, mode

- Mean: the sum of the values of a variable and dividing by number of the observations

Summary statistics:

Measures of Central Tendency: mean, median, mode

- Mean: the sum of the values of a variable and dividing by number of the observations
- Median: the middle (the 50th centile) observation

Summary statistics:

Measures of Central Tendency: mean, median, mode

- Mean: the sum of the values of a variable and dividing by number of the observations
- Median: the middle (the 50th centile) observation
- Mode: the value that occurs most frequently

Summary statistics:

Measures of Central Tendency: mean, median, mode

- Mean: the sum of the values of a variable and dividing by number of the observations
- Median: the middle (the 50th centile) observation
- Mode: the value that occurs most frequently

```
. tabstat age // will only give you mean
```

variable	mean
age	56.41176

```
. tabstat age, s(count mean median) // s stands for statistics
```

variable	N	mean	p50
age	34	56.41176	56

Summary statistics: Measures of Dispersion: range, IQR, variance, standard deviation

- Range: the difference between the maximum and the minimum

Summary statistics: Measures of Dispersion: range, IQR, variance, standard deviation

- Range: the difference between the maximum and the minimum
- Interquartile range: the absolute difference between the 25th percentile of the observations and the 75th.

Summary statistics: Measures of Dispersion: range, IQR, variance, standard deviation

- Range: the difference between the maximum and the minimum
- Interquartile range: the absolute difference between the 25th percentile of the observations and the 75th.
- Variance, standard deviation (sd): a measure of spread of the data

Summary statistics: Measures of Dispersion: range, IQR, variance, standard deviation

- Range: the difference between the maximum and the minimum
- Interquartile range: the absolute difference between the 25th percentile of the observations and the 75th.
- Variance, standard deviation (sd): a measure of spread of the data

$$s^2 = \widehat{Var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

```
. tabstat age, s(count range min max iqr var sd)
```

variable	N	range	min	max
-----+-----				
age	34	20	47	67
-----+-----				

variable	iqr	variance	sd
-----+-----			
age	10	36.12834	6.010686
-----+-----			

Tables: Bad example

What is the problem here?

Table 5
Simulation results for using full data, CRs only, and proposed method under four missing mechanisms

Method	Bias ^a		Variance ^b		95% CI ^c	
	$(\hat{\beta}_W)$	$(\hat{\beta}_X)$	$(\hat{\beta}_W)$	$(\hat{\beta}_X)$	$(\hat{\beta}_W)$	$(\hat{\beta}_X)$
(M.1) $P(R = 1) = 0.66$						
Full	0.01346	0.02229	0.04008	0.03685	0.955	0.950
Comp	0.03062	-0.003561	0.1149	0.06732	0.960	0.955
Impu	0.01431	0.021	0.04088	0.05169	0.980	0.975
(M.2) $\text{logit } P(R = 1) = 2Y$						
Full	0.007908	-0.02116	0.03838	0.03624	0.975	0.925
Comp	0.01945	0.07096	0.107	0.06581	0.960	0.950
Impu	0.006966	0.01597	0.04227	0.05226	0.975	0.985
(M.3) $\text{logit } P(R = 1) = 2X$						
Full	0.007908	-0.02116	0.03838	0.03624	0.975	0.925
Comp	0.01225	0.0589	0.08856	0.06818	0.980	0.975
Impu	0.009563	-0.04699	0.03865	0.04923	0.985	0.970
(M.4) $\text{logit } P(R = 1) = X + Y$						
Full	0.01346	0.02229	0.04008	0.03685	0.955	0.950
Comp	0.02404	1.613	0.1102	0.08202	0.955	0.580
Impu	0.01814	0.08289	0.0578	0.06075	0.955	0.970

^aBias = $(\hat{\beta} - \beta_0)/\beta_0$.

^bSimulation variance.

^cConfidence interval using jackknife standard error.

Tables: Basics of making tables

- Headings should be
 - self-explanatory and informative
 - placed above the tables.
 - (Graph headings are placed below.)

Tables: Basics of making tables

- Headings should be
 - self-explanatory and informative
 - placed above the tables.
 - (Graph headings are placed below.)
- Tables are numbered sequentially.

Tables: Basics of making tables

- Headings should be
 - self-explanatory and informative
 - placed above the tables.
 - (Graph headings are placed below.)
- Tables are numbered sequentially.
- Don't insert tables from the program's output.

Tables: Basics of making tables

- Headings should be
 - self-explanatory and informative
 - placed above the tables.
 - (Graph headings are placed below.)
- Tables are numbered sequentially.
- Don't insert tables from the program's output.
- No vertical lines
- As few horizontal lines as possible

Tables: Basics of making tables

- Headings should be
 - self-explanatory and informative
 - placed above the tables.
 - (Graph headings are placed below.)
- Tables are numbered sequentially.
- Don't insert tables from the program's output.
- No vertical lines
- As few horizontal lines as possible
- (Suggested) Arial 10 pt, normal spacing

Tables: Basics of making tables

- Headings should be
 - self-explanatory and informative
 - placed above the tables.
 - (Graph headings are placed below.)
- Tables are numbered sequentially.
- Don't insert tables from the program's output.
- No vertical lines
- As few horizontal lines as possible
- (Suggested) Arial 10 pt, normal spacing

Table 1: Baseline characteristics of colon cancer patients diagnosed during 1981-1990, Sweden.

		<50	50-59	60-69	70-79	≥80	All
Colon cancer							
1981-1990	Patient size (n)	1 148	2 485	6 227	9 381	5 442	24 683
	Female (%)	51.74	52.43	50.09	51.61	59.28	53.01
	Proportion of censoring ¹ (%)	34.41	15.69	3.12	0.22	0.09	4.07

Tables: One-way tables

- We use cancer data still.
sysuse cancer, clear
keep if drug ==1 | drug == 2

Tables: One-way tables

- We use cancer data still.
sysuse cancer, clear
keep if drug ==1 | drug == 2
- One-way table of frequencies with mean and sd of age
. table died, contents(freq mean age sd age)

1 if				
patient				
died		Freq.	mean(age)	sd(age)
-----+-----				
0		9	55.1111	5.487359
1		25	56.88	6.227091

Tables: One-way tables

- One-way table of frequencies

```
. tabulate died
```

1 if			
patient			
died	Freq.	Percent	Cum.
-----+-----			
0	9	26.47	26.47
1	25	73.53	100.00
-----+-----			
Total	34	100.00	

Tables: One-way tables

- Create table I of baseline characteristics using `table1_mc`
- This command is useful. Play it on your own!
- See `help table1_mc`

```
. ssc install table1_mc, replace  
. table1_mc, vars(age conts)
```

+-----+	
	Total
+-----+	
	N=34
+-----+	
Patient's age at start of exp.	56 (51-61)
+-----+	

Data are presented as median (IQR).

Tables: Two by two tables

- 2 by 2 table for drug and died with relative frequency by column or row

```
. tabulate died drug, col row
```

1 if patient Drug type (1=placebo) died	0		1	Total
0	8	1		9
	88.89	11.11		100.00
	57.14	5.00		26.47
1	6	19		25
	24.00	76.00		100.00
	42.86	95.00		73.53
Total	14	20		34
	41.18	58.82		100.00
	100.00	100.00		100.00

Tables: Two by two tables

- 2 by 2 table with chi-square test and fisher's exact test

```
. tabulate died drug, col row chi2 exact
```

1 if			
patient Drug type (1=placebo)			
died	0	1	Total
-----+-----			
0	8	1	9
	88.89	11.11	100.00
	57.14	5.00	26.47
-----+-----			
1	6	19	25
	24.00	76.00	100.00
	42.86	95.00	73.53
-----+-----			
Total	14	20	34
	41.18	58.82	100.00
	100.00	100.00	100.00

```
Pearson chi2(1) = 11.5039    Pr = 0.001
Fisher's exact =              0.001
```


Tables: Two by two tables

- 2 by 2 table with chi-square test and fisher's exact test

- How to interpret the results?

```
. tabulate died drug, col row chi2 exact
```

1 if			
patient Drug type (1=placebo)			
died	0	1	Total
-----+-----			
0	8	1	9
	88.89	11.11	100.00
	57.14	5.00	26.47
-----+-----			
1	6	19	25
	24.00	76.00	100.00
	42.86	95.00	73.53
-----+-----			
Total	14	20	34
	41.18	58.82	100.00
	100.00	100.00	100.00

```
Pearson chi2(1) = 11.5039 Pr = 0.001
Fisher's exact = 0.001
```

Tables: Two by two tables

- 2 by 2 table with chi-square test and fisher's exact test

```
. tabulate died drug, col row chi2 exact
1 if |
patient | Drug type (1=placebo)
died |      0      1 |      Total
-----+-----+-----
0 |      8      1 |      9
  |  88.89  11.11 |  100.00
  |  57.14   5.00 |  26.47
-----+-----+-----
1 |      6     19 |     25
  |  24.00  76.00 |  100.00
  |  42.86  95.00 |  73.53
-----+-----+-----
Total |     14     20 |     34
  |  41.18  58.82 |  100.00
  | 100.00 100.00 |  100.00
```

```
Pearson chi2(1) = 11.5039   Pr = 0.001
Fisher's exact =           0.001
```

- How to interpret the results?
- Chi-square test: testing the association between two binary variables.

Tables: Two by two tables

- 2 by 2 table with chi-square test and fisher's exact test

```
. tabulate died drug, col row chi2 exact
1 if |
patient | Drug type (1=placebo)
died |      0      1 |      Total
-----+-----+-----
0 |      8      1 |      9
  |  88.89  11.11 |  100.00
  |  57.14   5.00 |  26.47
-----+-----+-----
1 |      6     19 |     25
  |  24.00  76.00 |  100.00
  |  42.86  95.00 |  73.53
-----+-----+-----
Total |     14     20 |     34
  |  41.18  58.82 |  100.00
  | 100.00 100.00 |  100.00

Pearson chi2(1) = 11.5039   Pr = 0.001
Fisher's exact =           0.001
```

- How to interpret the results?
- Chi-square test: testing the association between two binary variables.
- Using placebo has association with that the patients died or not.

Tables: Two by two tables

- 2 by 2 tables stratified by sex

```
. bysort sex: tab died drug, col row chi2
```

```
-> sex = 0
```

1 if patient	Drug type (1=placebo)		
died	0	1	Total

0	6	1	7
	85.71	14.29	100.00
	75.00	9.09	36.84

1	2	10	12
	16.67	83.33	100.00
	25.00	90.91	63.16

Total	8	11	19
	42.11	57.89	100.00
	100.00	100.00	100.00

```
Pearson chi2(1) = 8.6466 Pr = 0.003
```

```
-> sex = 1
```

1 if patient	Drug type (1=placebo)		
died	0	1	Total

0	2	0	2
	100.00	0.00	100.00
	33.33	0.00	13.33

1	4	9	13
	30.77	69.23	100.00
	66.67	100.00	86.67

Total	6	9	15
	40.00	60.00	100.00
	100.00	100.00	100.00

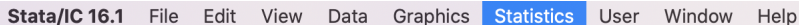
```
Pearson chi2(1) = 3.4615 Pr = 0.063
```

Tables: Stata tool for Epidemiology

- How to use Stata to generate risk ratios and odds ratios?

Tables: Stata tool for Epidemiology

- How to use Stata to generate risk ratios and odds ratios?
- A useful tool in Stata's default function can be found at
- Statistics - Epidemiology and related - Tables for epidemiologists

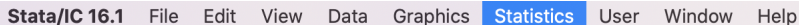


The image shows a horizontal menu bar from the Stata software interface. The menu items are: Stata/IC 16.1, File, Edit, View, Data, Graphics, Statistics, User, Window, and Help. The 'Statistics' menu item is highlighted with a blue background, while the others have a light gray background.

Stata/IC 16.1	File	Edit	View	Data	Graphics	Statistics	User	Window	Help
---------------	------	------	------	------	----------	------------	------	--------	------

Tables: Stata tool for Epidemiology

- How to use Stata to generate risk ratios and odds ratios?
- A useful tool in Stata's default function can be found at
- Statistics - Epidemiology and related - Tables for epidemiologists



The image shows a horizontal menu bar from the Stata software interface. It includes the text 'Stata/IC 16.1' followed by several menu items: 'File', 'Edit', 'View', 'Data', 'Graphics', 'Statistics', 'User', 'Window', and 'Help'. The 'Statistics' menu item is highlighted with a blue background, while the others have a light gray background.

- But before demonstrating how this works, a recapture on basic epi terms!

Basic Epidemiology terms: Rate vs. proportion

Rate

Proportion

Basic Epidemiology terms: Rate vs. proportion

Rate

- Incidence (rate): $\frac{\text{no. of diseased}}{\text{total person-time}}$

Proportion

Basic Epidemiology terms: Rate vs. proportion

Rate

- Incidence (rate): $\frac{\text{no. of diseased}}{\text{total person-time}}$
- Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$

Proportion

Basic Epidemiology terms: Rate vs. proportion

Rate

- Incidence (rate): $\frac{\text{no. of diseased}}{\text{total person-time}}$
- Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$
- Hazard (rate): in survival analysis, hazard is often defined as mortality rate.

Proportion

Basic Epidemiology terms: Rate vs. proportion

Rate

- Incidence (rate): $\frac{\text{no. of diseased}}{\text{total person-time}}$
- Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$
- Hazard (rate): in survival analysis, hazard is often defined as mortality rate.

Proportion

- Cumulative incidence: **over a time period**, $\frac{\text{no. of new cases of the disease}}{\text{no. of initially disease-free persons}}$

Basic Epidemiology terms: Rate vs. proportion

Rate

- Incidence (rate): $\frac{\text{no. of diseased}}{\text{total person-time}}$
- Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$
- Hazard (rate): in survival analysis, hazard is often defined as mortality rate.

Proportion

- Cumulative incidence: **over a time period**, $\frac{\text{no. of new cases of the disease}}{\text{no. of initially disease-free persons}}$
- Fatality: **over a time period**, $\frac{\text{no. of deaths of the disease}}{\text{no. of persons with the disease}}$

Basic Epidemiology terms: Rate vs. proportion

Rate

- Incidence (rate): $\frac{\text{no. of diseased}}{\text{total person-time}}$
- Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$
- Hazard (rate): in survival analysis, hazard is often defined as mortality rate.

Proportion

- Cumulative incidence: **over a time period**, $\frac{\text{no. of new cases of the disease}}{\text{no. of initially disease-free persons}}$
- Fatality: **over a time period**, $\frac{\text{no. of deaths of the disease}}{\text{no. of persons with the disease}}$
- Point prevalence: **at a specified time**, $\frac{\text{no. of diseased}}{\text{no. of persons}}$
- Period prevalence: **over a time period**, $\frac{\text{no. of diseased}}{\text{no. of persons}}$

Basic Epidemiology terms: Rate vs. proportion

Rate

- Incidence (rate): $\frac{\text{no. of diseased}}{\text{total person-time}}$
- Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$
- Hazard (rate): in survival analysis, hazard is often defined as mortality rate.

Proportion

- Cumulative incidence: **over a time period**, $\frac{\text{no. of new cases of the disease}}{\text{no. of initially disease-free persons}}$
- Fatality: **over a time period**, $\frac{\text{no. of deaths of the disease}}{\text{no. of persons with the disease}}$
- Point prevalence: **at a specified time**, $\frac{\text{no. of diseased}}{\text{no. of persons}}$
- Period prevalence: **over a time period**, $\frac{\text{no. of diseased}}{\text{no. of persons}}$
- Survival (proportion/probability) ~~rate~~:
$$\frac{\text{no. of alive persons (since diagnosis)}}{\text{no. of initially disease-free persons (since diagnosis)}}$$

Basic Epidemiology terms: Rate vs. proportion

Quizzes

Basic Epidemiology terms: Rate vs. proportion

Quizzes

1. What is the key difference between rate and proportion?

Quizzes

1. What is the key difference between rate and proportion?
 - TIME!
 - Rate: person-time
 - Proportion: specify a period/point of time

Quizzes

1. What is the key difference between rate and proportion?
 - TIME!
 - Rate: person-time
 - Proportion: specify a period/point of time
2. Mortality vs. fatality?

Basic Epidemiology terms: Rate vs. proportion

Quizzes

1. What is the key difference between rate and proportion?

- TIME!
- Rate: person-time
- Proportion: specify a period/point of time

2. Mortality vs. fatality?

- Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$
- Fatality: over a time period, $\frac{\text{no. of deaths of the disease}}{\text{no. of persons with the disease}}$

Basic Epidemiology terms: Rate vs. proportion

Quizzes

1. What is the key difference between rate and proportion?
 - TIME!
 - Rate: person-time
 - Proportion: specify a period/point of time
2. Mortality vs. fatality?
 - Mortality (rate): $\frac{\text{no. of deaths}}{\text{total person-time}}$
 - Fatality: over a time period, $\frac{\text{no. of deaths of the disease}}{\text{no. of persons with the disease}}$
3. Is risk a rate or a proportion?

Basic Epidemiology terms: Risk, risk difference, risk ratio

Risk: the proportion (probability) of an event, e.g., death.

Basic Epidemiology terms: Risk, risk difference, risk ratio

Risk: the proportion (probability) of an event, e.g., death.

- Estimates of risk: cumulative incidence, cumulative hazard

Basic Epidemiology terms: Risk, risk difference, risk ratio

Risk: the proportion (probability) of an event, e.g., death.

- Estimates of risk: cumulative incidence, cumulative hazard
- E.g., in survival analysis,

Cumulative hazard = $1 - \text{Survival proportion}$ = Cumulative probability of death

$$F(t) = 1 - S(t) = P(T \leq t)$$

Basic Epidemiology terms: Risk, risk difference, risk ratio

Risk: the proportion (probability) of an event, e.g., death.

- Estimates of risk: cumulative incidence, cumulative hazard
- E.g., in survival analysis,

Cumulative hazard = $1 - \text{Survival proportion}$ = Cumulative probability of death

$$F(t) = 1 - S(t) = P(T \leq t)$$

- Risk difference: the difference of the probabilities of an event between the exposed group and non-exposed group

Basic Epidemiology terms: Risk, risk difference, risk ratio

Risk: the proportion (probability) of an event, e.g., death.

- Estimates of risk: cumulative incidence, cumulative hazard
- E.g., in survival analysis,

Cumulative hazard = $1 - \text{Survival proportion}$ = Cumulative probability of death

$$F(t) = 1 - S(t) = P(T \leq t)$$

- Risk difference: the difference of the probabilities of an event between the exposed group and non-exposed group
- Risk ratio: the ratio of the probabilities of an event between the exposed group and non-exposed group

Basic Epidemiology terms: Risk, risk difference, risk ratio

Risk: the proportion (probability) of an event, e.g., death.

- Estimates of risk: cumulative incidence, cumulative hazard
- E.g., in survival analysis,

Cumulative hazard = $1 - \text{Survival proportion}$ = Cumulative probability of death

$$F(t) = 1 - S(t) = P(T \leq t)$$

- Risk difference: the difference of the probabilities of an event between the exposed group and non-exposed group
- Risk ratio: the ratio of the probabilities of an event between the exposed group and non-exposed group
- **Caution!** Relative risk could be either risk ratio or rate ratio!

Basic Epidemiology terms: Risk, risk difference, risk ratio

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

Epidemiologists love two by two tables!

Basic Epidemiology terms: Risk, risk difference, risk ratio

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

Epidemiologists love two by two tables!

- Risk difference between females having shiba and males having shiba = $\widehat{p}_F - \widehat{p}_M = 2/4 - 2/3 = -0.16667$
- Interpretation: Females have 16.67 % lower risk of having shiba than males.

Basic Epidemiology terms: Risk, risk difference, risk ratio

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

Epidemiologists love two by two tables!

- Risk difference between females having shiba and males having shiba
$$= \widehat{p}_F - \widehat{p}_M = 2/4 - 2/3 = -0.16667$$
- Interpretation: Females have 16.67 % lower risk of having shiba than males.
- Risk ratio between females having shiba and males having shiba
$$= \widehat{p}_F \div \widehat{p}_M = 2/4 \div 2/3 = 0.75.$$
- Interpretation: The RR of females having shiba is 0.75 times than males having shiba.

Basic Epidemiology terms: Odds, odds ratio

Odds: the ratio between those having and not having an outcome.

$$Odds = \frac{p}{1 - p}$$

Basic Epidemiology terms: Odds, odds ratio

Odds: the ratio between those having and not having an outcome.

$$Odds = \frac{p}{1 - p}$$

- Odds ratio measures the association between an exposure and an outcome.

$$\text{Odds ratio (OR)} = \frac{Odds_{exposed}}{Odds_{unexposed}}$$

Basic Epidemiology terms: Odds, odds ratio

Odds: the ratio between those having and not having an outcome.

$$Odds = \frac{p}{1 - p}$$

- Odds ratio measures the association between an exposure and an outcome.

$$\text{Odds ratio (OR)} = \frac{Odds_{\text{exposed}}}{Odds_{\text{unexposed}}}$$

- E.g., the OR is 0.5, which indicates that there is a 50% decrease in the odds of having an outcome among the exposed compared to the unexposed.

Basic Epidemiology terms: Odds, odds ratio

Odds: the ratio between those having and not having an outcome.

$$Odds = \frac{p}{1 - p}$$

- Odds ratio measures the association between an exposure and an outcome.

$$\text{Odds ratio (OR)} = \frac{Odds_{\text{exposed}}}{Odds_{\text{unexposed}}}$$

- E.g., the OR is 0.5, which indicates that there is a 50% decrease in the odds of having an outcome among the exposed compared to the unexposed.
- Why there is no odds difference?

Basic Epidemiology terms: Odds, odds ratio

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

Basic Epidemiology terms: Odds, odds ratio

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

- The odds of having shiba among females is

$$\begin{aligned}\widehat{odds}_F &= \frac{p(\text{having shiba}|\text{female})}{p(\text{having guinea pig}|\text{female})} \\ &= \frac{(2/4)}{(2/4)} = 1\end{aligned}$$

- The odds of having shiba among males is 2 (calculation ignored).

Basic Epidemiology terms: Odds, odds ratio

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

- OR of having shiba (females to males)
- $OR = \frac{Odds_f}{Odds_m} = \frac{1}{2}$

- The odds of having shiba among females is

$$\begin{aligned}\widehat{odds}_f &= \frac{p(\text{having shiba}|\text{female})}{p(\text{having guinea pig}|\text{female})} \\ &= \frac{(2/4)}{(2/4)} = 1\end{aligned}$$

- The odds of having shiba among males is 2 (calculation ignored).

Basic Epidemiology terms: Odds, odds ratio

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

- The odds of having shiba among females is

$$\begin{aligned}\widehat{odds}_f &= \frac{p(\text{having shiba}|\text{female})}{p(\text{having guinea pig}|\text{female})} \\ &= \frac{(2/4)}{(2/4)} = 1\end{aligned}$$

- The odds of having shiba among males is 2 (calculation ignored).

- OR of having shiba (females to males)
- $OR = \frac{Odds_f}{Odds_m} = \frac{1}{2}$
- Interpretation: there is a 50% decrease in the odds of having shiba among females compared to males. Higher odds of shiba ownership among males than females!
- It seems that females instead love guinea pigs more.

Interpreting results: Principles

- When describing a ratio, it can ideally be illustrated by
 1. Exposed group
 2. Ratio (exact value, higher or lower percentage)
 3. Outcome
 4. Unexposed

Interpreting results: Principles

- When describing a ratio, it can ideally be illustrated by
 1. Exposed group
 2. Ratio (exact value, higher or lower percentage)
 3. Outcome
 4. Unexposed
- Example:
 1. Females have a RR of 0.75 having shiba compared to males.

Interpreting results: Principles

- When describing a ratio, it can ideally be illustrated by
 1. Exposed group
 2. Ratio (exact value, higher or lower percentage)
 3. Outcome
 4. Unexposed
- Example:
 1. Females have a RR of 0.75 having shiba compared to males.
 2. Females have a 50% decrease in the odds of having shiba compared to males.

Interpreting results: Ratio $>$ or $<$ 1

Ratio

Interpreting results: Ratio $>$ or $<$ 1

Ratio

- As ratio $<$ 1,
 - $(1 - \text{RR/OR}) \times 100\%$
 - E.g., $\text{RR} = 0.75$, $(1 - 0.75) \times 100\% = 25\%$
 - 25% lower risk

Interpreting results: Ratio $>$ or $<$ 1

Ratio

- As ratio $<$ 1,
 - $(1 - \text{RR/OR}) \times 100\%$
 - E.g., $\text{RR} = 0.75$, $(1 - 0.75) \times 100\% = 25\%$
 - 25% lower risk
- As ratio $>$ 1,
 - $(\text{RR/OR} - 1) \times 100\%$
 - E.g., $\text{OR} = 2.05$, $(2.05 - 1) \times 100\% = 105\%$
 - The odds is 2 times higher.
 - Twice the odds

Interpreting results: More examples

Diabetes Is a Risk Factor for Pulmonary Tuberculosis: A Case-Control Study from Mwanza, Tanzania (Faurholt-Jepsen, 2011)

	OR (95% C.I.)	OR (95% C.I.)	OR (95% C.I.)
		Model 1	Model 2
	Unadjusted	Adjusted for age, sex, socio-demography ²	Model 1 + AGP ³
HIV negative (n = 770)			
Glucose intolerance status ¹			
normal glucose tolerance	ref.	ref.	ref.
IFG/IGT	2.26 (1.50;3.41)	2.34 (1.52;3.61)	2.65 (1.00;7.06)
diabetes	2.15 (1.35;3.42)	2.14 (1.32;3.46)	4.23 (1.54;11.57)

Interpreting results: More examples

Diabetes Is a Risk Factor for Pulmonary Tuberculosis: A Case-Control Study from Mwanza, Tanzania (Faurholt-Jepsen, 2011)

	OR (95% C.I.)	OR (95% C.I.)	OR (95% C.I.)
		Model 1	Model 2
	Unadjusted	Adjusted for age, sex, socio-demography ²	Model 1 + AGP ³
HIV negative (n = 770)			
Glucose intolerance status ¹			
normal glucose tolerance	ref.	ref.	ref.
IFG/IGT	2.26 (1.50;3.41)	2.34 (1.52;3.61)	2.65 (1.00;7.06)
diabetes	2.15 (1.35;3.42)	2.14 (1.32;3.46)	4.23 (1.54;11.57)

1. People with diabetes had a higher odds of TB (OR 2.15, 95% CI: 1.35-3.42) relative to people without diabetes.

Interpreting results: More examples

Diabetes Is a Risk Factor for Pulmonary Tuberculosis: A Case-Control Study from Mwanza, Tanzania (Faurholt-Jepsen, 2011)

	OR (95% C.I.)	OR (95% C.I.)	OR (95% C.I.)
		Model 1	Model 2
	Unadjusted	Adjusted for age, sex, socio-demography ²	Model 1 + AGP ³
HIV negative (n = 770)			
Glucose intolerance status ¹			
normal glucose tolerance	ref.	ref.	ref.
IFG/IGT	2.26 (1.50;3.41)	2.34 (1.52;3.61)	2.65 (1.00;7.06)
diabetes	2.15 (1.35;3.42)	2.14 (1.32;3.46)	4.23 (1.54;11.57)

1. People with diabetes had a higher odds of TB (OR 2.15, 95% CI: 1.35-3.42) relative to people without diabetes.
2. Having diabetes was associated with more than a 2-fold increase (OR: 2.15, 95% CI: 1.35; 3.42) in the odds of TB compared to not having diabetes.

Interpreting results: More examples

Bidirectional association between physical activity and symptoms of anxiety and depression: the Whitehall II study (Azevedo Da Silva, 2012)

Table 3 Cross-sectional associations between physical activity at recommended levels and anxiety and/or depression symptoms at phase 1 (1985–1988) (N = 9,309)

	OR (CI 95 %)	<i>P</i> value
<i>Anxiety symptoms</i>		
Model 1		
Physical activity		
Yes	0.71 (0.54, 0.91)	0.01
No	1 (reference)	
Model 2		
Physical activity		
Yes	0.71 (0.55, 0.93)	0.01
No	1 (reference)	
<i>Depression symptoms</i>		
Model 1		
Physical activity		
Yes	0.63 (0.48, 0.81)	<0.001
No	1 (reference)	
Model 2		
Physical activity		
Yes	0.63 (0.49, 0.82)	0.001
No	1 (reference)	

Interpreting results: More examples

Bidirectional association between physical activity and symptoms of anxiety and depression: the Whitehall II study (Azevedo Da Silva, 2012)

Table 3 Cross-sectional associations between physical activity at recommended levels and anxiety and/or depression symptoms at phase 1 (1985–1988) (N = 9,309)

	OR (CI 95 %)	<i>P</i> value
<i>Anxiety symptoms</i>		
Model 1		
Physical activity		
Yes	0.71 (0.54, 0.91)	0.01
No	1 (reference)	
Model 2		
Physical activity		
Yes	0.71 (0.55, 0.93)	0.01
No	1 (reference)	
<i>Depression symptoms</i>		
Model 1		
Physical activity		
Yes	0.63 (0.48, 0.81)	<0.001
No	1 (reference)	
Model 2		
Physical activity		
Yes	0.63 (0.49, 0.82)	0.001
No	1 (reference)	

1. Patients who conducted recommended levels of physical activity had a 29% lower odds of anxiety (OR: 0.71, 95% CI: 0.54-0.91) and a 37% lower odds of depression (OR: 0.63, 95% CI: 0.48-0.81) relative to those who did not.

Interpreting results: More examples

Bidirectional association between physical activity and symptoms of anxiety and depression: the Whitehall II study (Azevedo Da Silva, 2012)

Table 3 Cross-sectional associations between physical activity at recommended levels and anxiety and/or depression symptoms at phase 1 (1985–1988) (N = 9,309)

	OR (CI 95 %)	<i>P</i> value
<i>Anxiety symptoms</i>		
Model 1		
Physical activity		
Yes	0.71 (0.54, 0.91)	0.01
No	1 (reference)	
Model 2		
Physical activity		
Yes	0.71 (0.55, 0.93)	0.01
No	1 (reference)	
<i>Depression symptoms</i>		
Model 1		
Physical activity		
Yes	0.63 (0.48, 0.81)	<0.001
No	1 (reference)	
Model 2		
Physical activity		
Yes	0.63 (0.49, 0.82)	0.001
No	1 (reference)	

1. Patients who conducted recommended levels of physical activity had a 29% lower odds of anxiety (OR: 0.71, 95% CI: 0.54-0.91) and a 37% lower odds of depression (OR: 0.63, 95% CI: 0.48-0.81) relative to those who did not.
2. Our results showed that individuals who practiced recommended levels of physical activity were less likely to have anxiety (OR: 0.71, 95% CI: 0.54-0.91) and depression (OR: 0.63, 95% CI: 0.48-0.81) in comparison with those who did not.

Calculate ratios using Stata: Risk ratio

- Finally we come back to Stata again!

Calculate ratios using Stata: Risk ratio

- Finally we come back to Stata again!
- cs case exposed

```
. cs died drug
```

	Drug type [1=placebo]		
	Exposed	Unexposed	Total
Cases	19	6	25
Noncases	1	8	9
Total	20	14	34
Risk	.95	.4285714	.7352941
	Point estimate		[95% Conf. Interval]
Risk difference	.5214286		.245166 .7976911
Risk ratio	2.216667		1.200631 4.092525
Attr. frac. ex.	.5488722		.1671043 .7556521
Attr. frac. pop	.4171429		
+-----			
chi2(1) = 11.50 Pr>chi2 = 0.0007			

Calculate ratios using Stata: Odds ratio

- cs case exposed, or
. cs died drug, or

	Drug type [1=placebo]			
	Exposed	Unexposed	Total	
Cases	19	6	25	
Noncases	1	8	9	
Total	20	14	34	
Risk	.95	.4285714	.7352941	
	Point estimate		[95% Conf. Interval]	
Risk difference	.5214286		.245166	.7976911
Risk ratio	2.216667		1.200631	4.092525
Attr. frac. ex.	.5488722		.1671043	.7556521
Attr. frac. pop	.4171429			
Odds ratio	25.33333		3.189793	. (Cornfield
+-----				

Calculate ratios using Stata: Incidence rate ratio

- ir case exposed studytime

```
. ir died drug studytime
```

```
Incidence-rate comparison
```

	Drug type [1=placebo]			
	Exposed	Unexposed	Total	
1 if patient die	19	6	25	
Months to death	180	209	389	
Incidence rate	.1055556	.0287081	.0642674	
	Point estimate		[95% Conf. Interval]	
Inc. rate diff.	.0768474		.0241182	.1295766
Inc. rate ratio	3.676852		1.411772	11.24864 (exact)
Attr. frac. ex.	.7280282		.2916701	.9111003 (exact)
Attr. frac. pop	.5533014			

```
Mid p-values for tests of incidence-rate difference:
```

```
Adj Pr(Exposed 1 if patient die <= 19) = 0.9985 (lower one-sided)
```

```
Adj Pr(Exposed 1 if patient die >= 19) = 0.0015 (upper one-sided)
```

```
Two-sided p-value = 0.0031
```

References

1. Gordis L. *Epidemiology*. Philadelphia, PA: Elsevier/Saunders, 2014. ISBN 9781455737338.
2. David Goldsman PG. *A First Course in Probability and Statistics*. Georgia Institute of Technology, 2020.
3. Marcello Pagano KG. *Principles of Biostatistics*. Cengage Learning, Inc, 2000. ISBN 0534229026.
4. Erin Gabriel PF. Epidemiology PhD program, Karolinska Institutet, 2020.