



SUMMARY STATISTICS

SYRACUSE UNIVERSITY
School of Information Studies

SUMMARY STATISTICS

Common summary statistics:

- Central tendency of data

- Data spread

Different statistical measures for different variable types:

- Numeric

- Nominal

NUMERIC VARIABLES

Central tendency

- Mean

- Median

- Mode

Data spread

- Standard deviation

- Variance

- Min., max., quartiles

DATA SPREAD

Range: Max. – Min.

Variance/standard deviation

Quartile: Q1 (25%), Q2 (50%), Q3 (75%), Q4 (100%)

E.g.,
temperature=[55,55,**Q1**56,58,60,**median**60,60,61,**Q3**70,72,74]

Interquartile range (IQR): $Q3 - Q1 = 70 - 56 = 14$

SUMMARIZE NUMERIC VARIABLE IN R

Summarize the whole data set:

```
summary(titanic)
```

Summarize central tendency:

```
mean(titanic$Age)
```

```
median(titanic$Age)
```

```
freq=table(titanic$Age)
```

```
table(titanic$Age)[which.max(table(titanic$Age))] # mode
```

SUMMARIZE NUMERIC VARIABLE IN R

Summarize data spread:

```
var(titanic$Age) # variance
```

```
sd(titanic$Age) # standard deviation
```

```
max(titanic$Age)
```

```
min(titanic$Age)
```

```
range <- max(titanic$Age) - min(titanic$Age)
```

```
qt <- quantile(titanic$Age, na.rm=TRUE) # quartile, remove  
missing values
```

```
IQR=qt[['75%']]-qt[['25%']] # Interquartile range
```

NOMINAL VARIABLES

Central tendency: Mode

Data spread: Distribution

```
> table(titanic$Sex)

female    male
   314     577

> table(titanic$Sex)[which.max(table(titanic$Sex))]
male
577
```

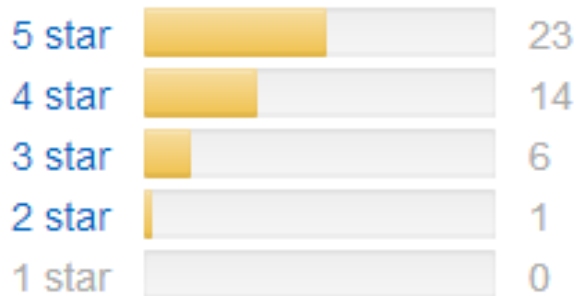
ORDINAL VARIABLES

Can we treat ordinal variables as nominal or numeric?

Customer Reviews

★★★★★ (44)

4.3 out of 5 stars



[See all 44 customer reviews ▶](#)

CUSTOMER ID	RATING
1	*
2	*****
3	**
4	***
5	****
6	**
...	...
Average	?