



UNIVERSITAS
GADJAH MADA

Statistics and Probability

Department of Civil and Environmental Engineering

Faculty of Engineering

Universitas Gadjah Mada

2	1- Mampu memahami konsep dasar statistika, probabilitas, dan jenis-jenis distribusi probabilitas (IK a-1, CPMK-1)	<ul style="list-style-type: none">• Dapat menjelaskan dengan benar pengertian data, sampel dan populasi• Dapat mempresentasikan data dalam bentuk tabel atau grafik	Pengertian data, sampel, populasi dan presentasi data dalam bentuk tabel atau grafik	<ol style="list-style-type: none">a. Pengertian data, sampel, populasib. Presentasi data dalam bentuk tabel atau grafik
---	---	--	--	--



What is Data?





105-2686834-3759466	Ohio	Nashville	Kansas	10.99	440	N	B00000I5Y6	Katherine H.
105-9318443-4200264	Illinois	Orange County	Boston	16.99	312	Y	B000002BK9	Samuel P.
105-1872500-0198646	Massachusetts	Bad Blood	Chicago	15.98	413	N	B000068ZVQ	Chris G.
103-2628345-9238664	Canada	Let Go	Mammals	11.99	902	N	B0000010AA	Monique D.
002-1663369-6638649	Ohio	Best of Kansas	Kansas	10.99	440	N	B002MXA7Q0	Katherine H.



Order Number	Name	State/Country	Price	Area Code	Previous Album Download	Gift?	ASIN	New Purchase Artist
105-2686834-3759466	Katherine H.	Ohio	10.99	440	Nashville	N	B0000015Y6	Kansas
105-9318443-4200264	Samuel R	Illinois	16.99	312	Orange County	Y	B000002BK9	Boston
105-1372500-0198646	Chris G.	Massachusetts	15.98	413	Bad Blood	N	B000068ZVQ	Chicago
103-2628345-9238664	Monique D.	Canada	11.99	902	Let Go	N	B0000010AA	Mammals
002-1663369-6638649	Katherine H.	Ohio	10.99	440	Best of Kansas	N	B002MXA7Q0	Kansas

Any collection of numbers, characters, images, or other items that provide information about something

The discipline of statistics provides methods for organizing and summarizing data and for drawing conclusions based on information contained in the data.



Who and What

How the Data Are Collected

Categorical Variables



For each description of data, identify the W's, name the variables, specify for each variable whether its use indicates that it should be treated as categorical or quantitative, and, for any quantitative variable, identify the units in which it was measured (or note that they were not provided).

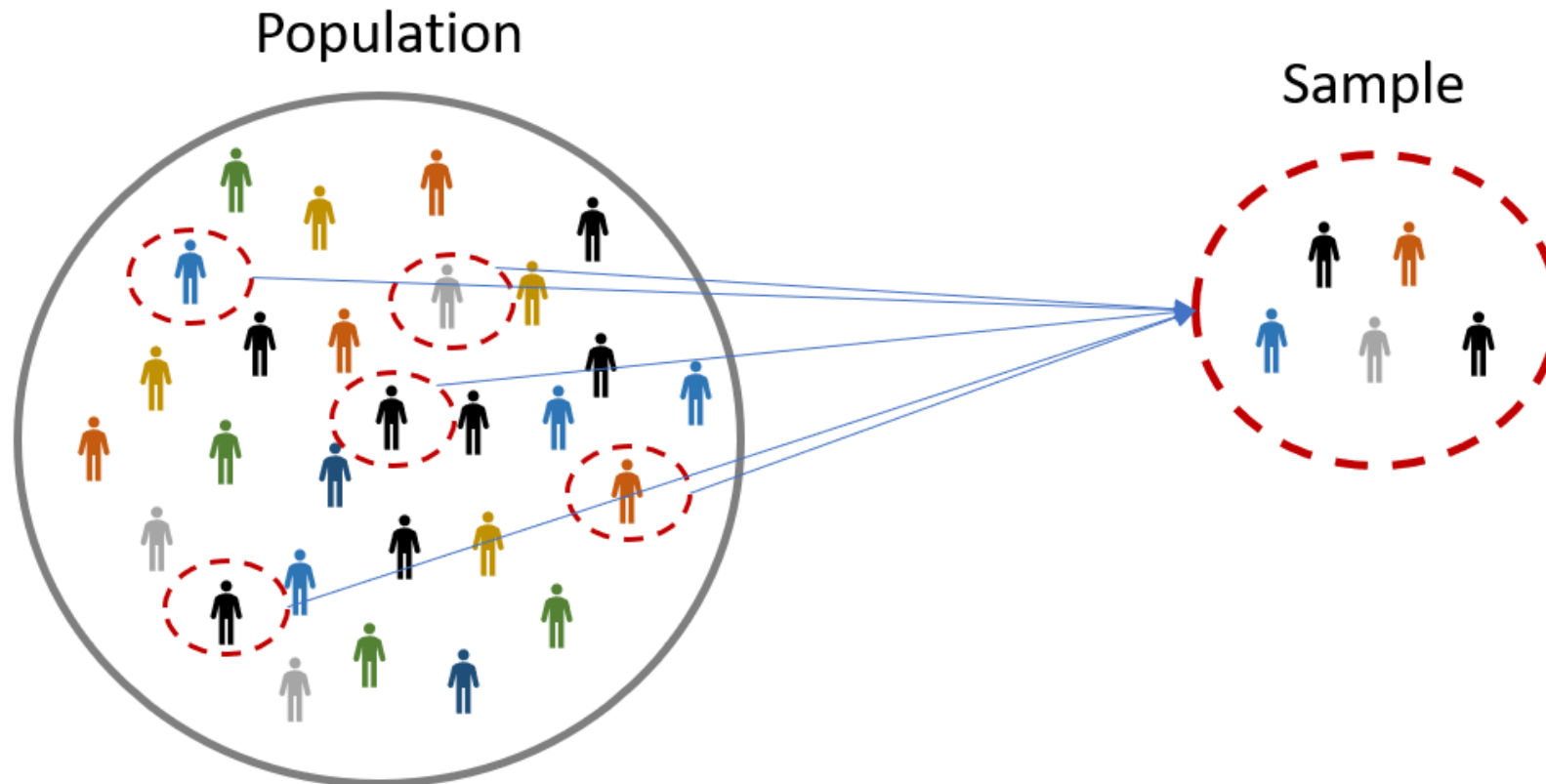
In 2006, *Consumer Reports* published an article evaluating refrigerators. It listed 41 models, giving the brand, cost, size (cu ft), type (such as top freezer), estimated annual energy cost, an overall rating (good, excellent, etc.), and the repair history for that brand (percentage requiring repairs over the past 5 years).

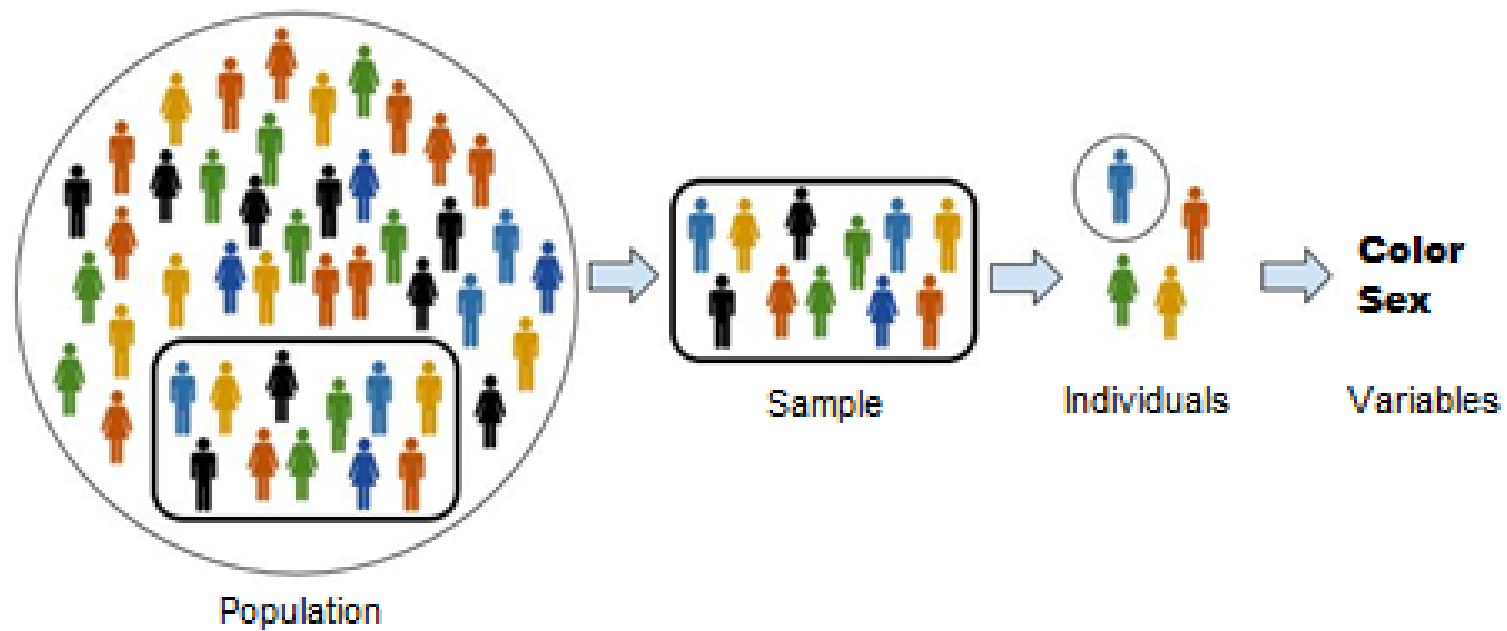


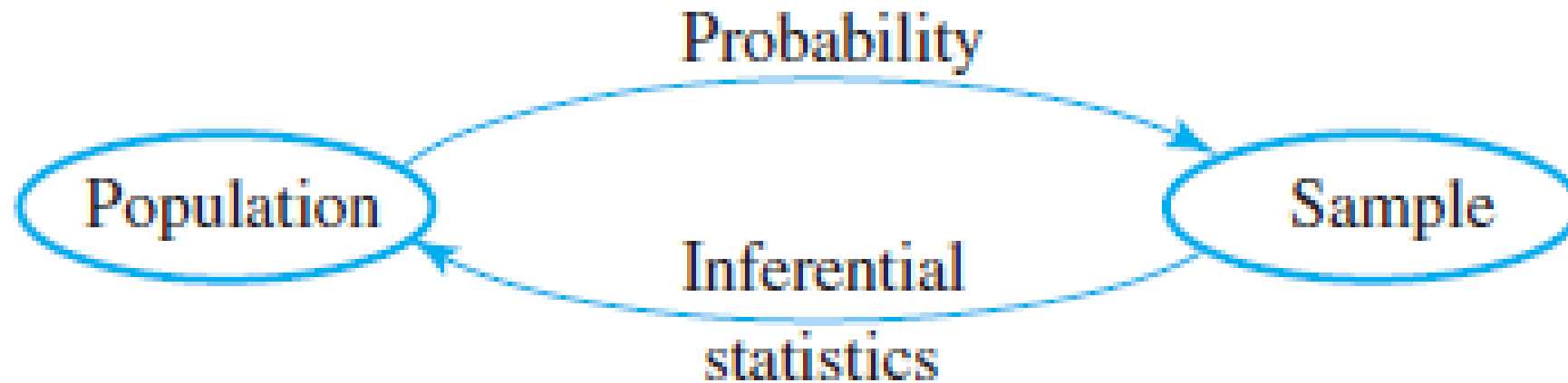
- Who : Refrigerator models
- What history : Brand, Cost, Types, Size, Overall rating, energy cost, repair
- When : 2006
- Where : United States
- Why : to provide information for consumer reports readers
- How : not specified



- Variable
- Cost Quantitative
- Size Quantitative
- Energy cost Quantitative
- Type Qualitative
- Repair history Quantitative
- Overall rating Qualitative
- Brand Qualitative







Inference

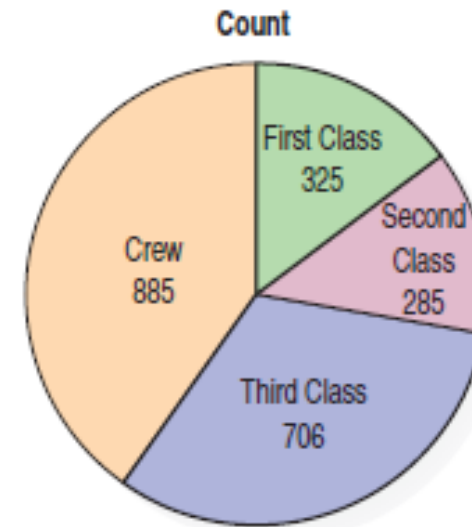
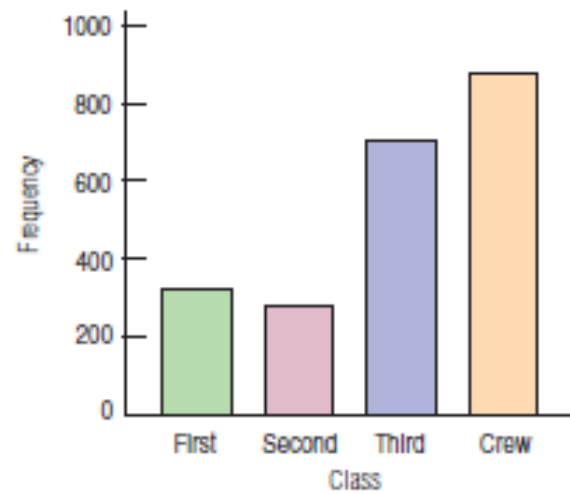
Process of making generalizations about a population's characteristics from the evidence in one sample

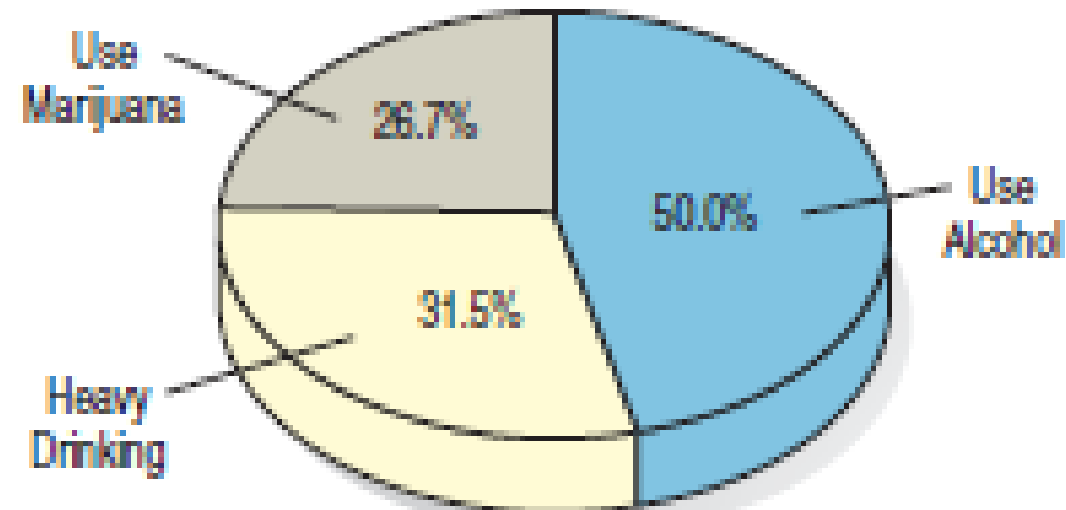
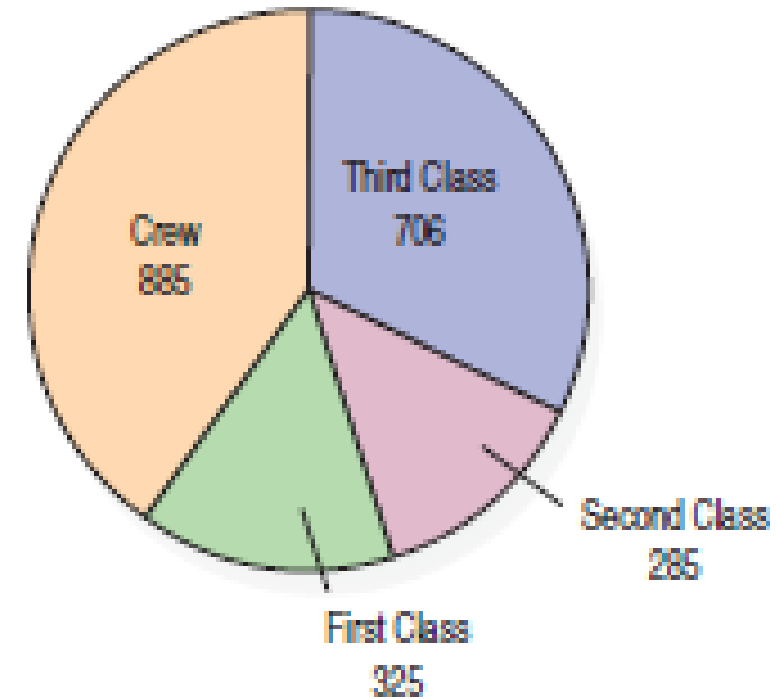
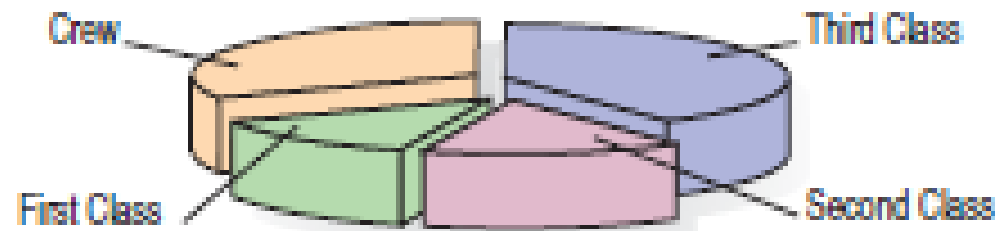
Valid inferences: representative sample using simple random sampling



The Three Rules of Data Analysis

1. Make a picture. A display of your data will reveal things you're not likely to see in a table of numbers and will help you to *Think* clearly about the patterns and relationships that may be hiding in your data.
2. Make a picture. A well-designed display will *Show* the important features and patterns in your data. It could also show you things you did not expect to see: extraordinary (possibly wrong) data values or unexpected patterns.
3. Make a picture. The best way to *Tell* others about your data is with a well-chosen picture.





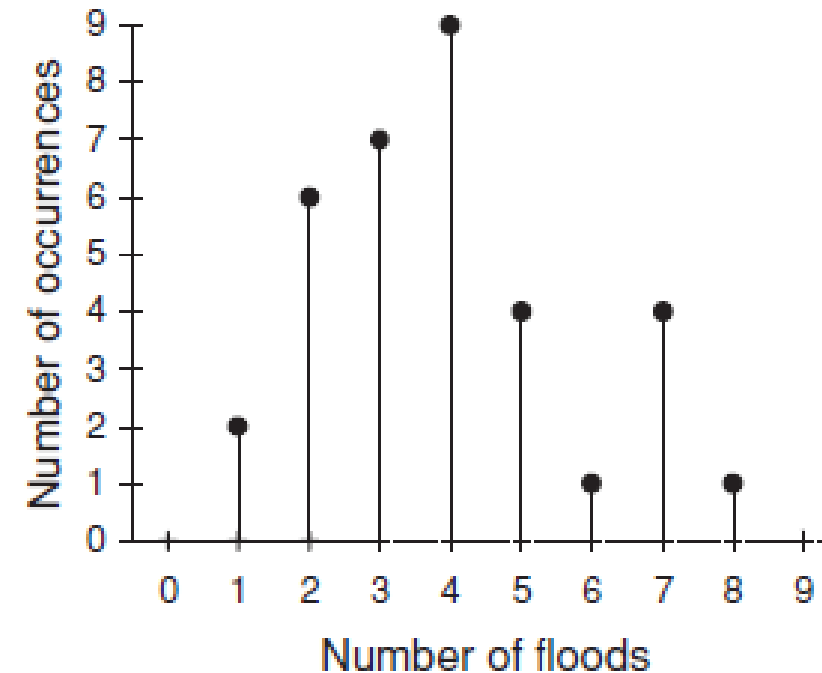
Graphical Representation



- Line Diagram or Bar Chart

Number of floods in a year	Number of occurrences
0	0
1	2
2	6
3	7
4	9
5	4
6	1
7	4
8	1
9	0
Total	34

^a A flood occurrence is defined as river discharge exceeding 300 m³/s.



Histogram



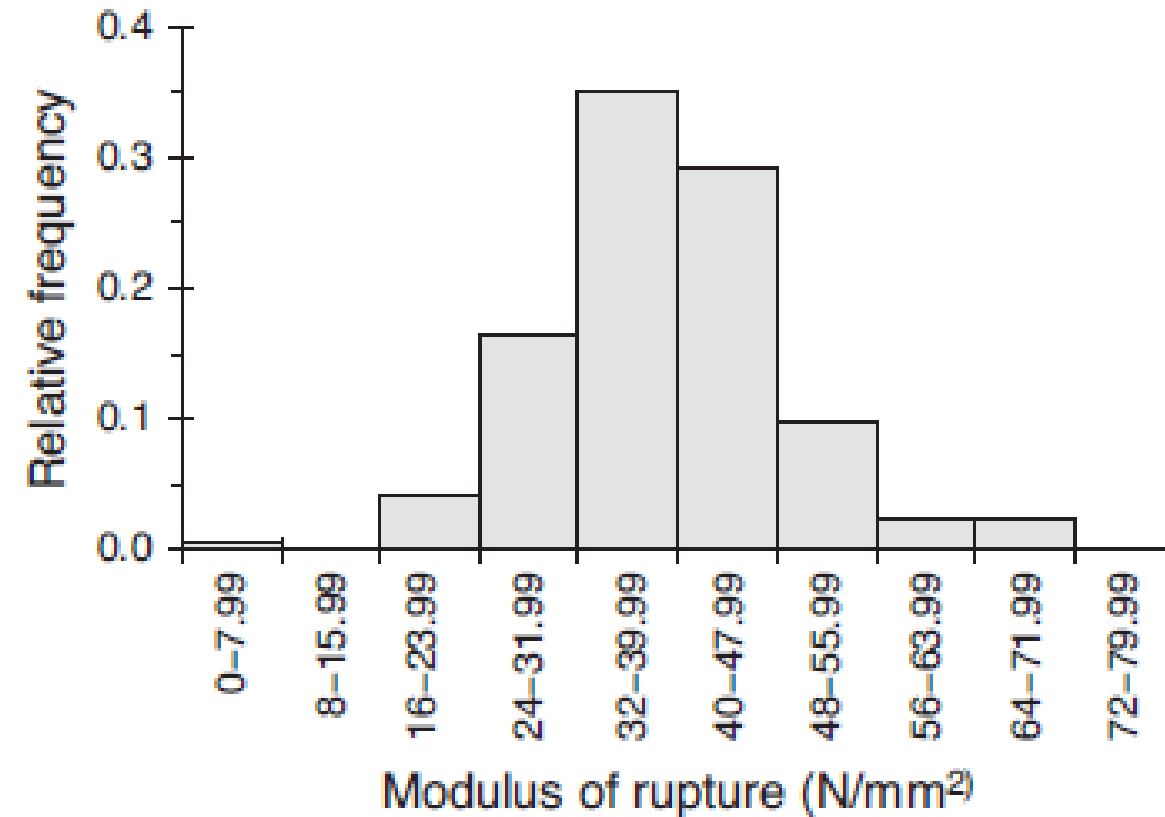
UNIVERSITAS GADJAH MADA

0.00	28.00	31.60	34.44	36.84	39.21	41.75	44.30	47.25	53.99
17.98	28.13	32.02	34.49	36.85	39.33	41.78	44.36	47.42	54.04
22.67	28.46	32.03	34.56	36.88	39.34	41.85	44.36	47.61	54.71
22.74	28.69	32.40	34.63	36.92	39.60	42.31	44.51	47.74	55.23
22.75	28.71	32.48	35.03	37.51	39.62	42.47	44.54	47.83	56.60
23.14	28.76	32.68	35.17	37.65	39.77	43.07	44.59	48.37	56.80
23.16	28.83	32.76	35.30	37.69	39.93	43.12	44.78	48.39	57.99
23.19	28.97	33.06	35.43	37.78	39.97	43.26	44.78	48.78	58.34
24.09	28.98	33.14	35.58	38.00	40.20	43.33	45.19	49.57	65.35
24.25	29.11	33.18	35.67	38.05	40.27	43.33	45.54	49.59	65.61
24.84	29.90	33.19	35.88	38.16	40.39	43.41	45.92	49.65	69.07
25.39	29.93	33.47	35.89	38.64	40.53	43.48	45.97	50.91	70.22
25.98	30.02	33.61	36.00	38.71	40.71	43.48	46.01	50.98	
26.63	30.05	33.71	36.38	38.81	40.85	43.64	46.33	51.39	
27.31	30.33	33.92	36.47	<u>39.05</u>	40.85	43.99	46.50	51.90	
27.90	30.53	34.12	36.53	39.15	41.64	44.00	46.86	53.00	
27.93	31.33	34.40	36.81	39.20	41.72	44.07	46.99	53.63	

^a The original data set is given in Table E.1.1; $n = 165$. The median is underlined.

Class upper limit (N/mm ²)	Class center (N/mm ²)	Absolute frequency	Relative frequency	Cumulative relative frequency (%)
5	2.5	1	0.006	0.61
10	7.5	0	0.000	0.61
15	12.5	0	0.000	0.61
20	17.5	1	0.006	1.21
25	22.5	9	0.055	6.67
30	27.5	18	0.109	17.58
35	32.5	26	0.158	33.33
40	37.5	38	0.230	56.36
45	42.5	34	0.206	76.97
50	47.5	20	0.121	89.09
55	52.5	9	0.055	94.55
60	57.5	5	0.030	97.58
65	62.5	0	0.000	97.58
70	67.5	3	0.018	99.39
75	72.5	1	0.006	100.00

^a The width of each class is 5 N/mm² in this example.



LOCALLY ROOTED, GLOBALLY RESPECTED

Drivers' licenses 2011 The table on the next page shows the number of licensed U.S. drivers (in millions) by age and by sex (www.dot.gov).

- What percent of total drivers are under 20?
- What percent of total drivers are male?
- Write a few sentences comparing the number of male and female licensed drivers in each age group.
- Do a driver's age and sex appear to be independent? Explain.

Age	Male Drivers (millions)	Female Drivers (millions)	Total
19 and Under	5.1	4.9	10.0
20–24	8.7	8.6	17.3
25–29	9.2	9.2	18.4
30–34	8.9	8.9	17.8
35–39	9.7	9.6	19.3
40–44	9.9	9.8	19.7
45–49	10.6	10.7	21.3
50–54	10.1	10.2	20.3
55–59	8.7	8.9	17.6
60–64	7.2	7.3	14.5
65–69	5.3	5.4	10.7
70–74	3.8	4.0	7.8
75–79	2.9	3.2	6.1
80–84	2.0	2.4	4.4
85 and Over	1.4	1.7	3.1
Total	103.5	104.8	208.3



- a) 4.8%
- b) 49.7%
- c) The drivers of each sex in each group are somewhat similar (there are about 50%)
Below age 25 more male drivers; above 45 more female drivers
- d) As the age increases, there is a small increase in the percentage of female drivers (slight correlations)