UCLA STAT 13

Introduction to Statistical Methods for the Life and Health Sciences

- **Instructor**: **Ivo Dinov**,
  Asst. Prof. In Statistics and Neurology
- **Teaching Assistants:** Hui Wang & Ming Zheng

University of California, Los Angeles, Fall 2004

http://www.stat.ucla.edu/~dinov/courses_students.html

UCLA STAT 13

to just hear is to forget

to see is to remember

to do it yourself is to understand …

(… to go to class is to … comprehend …)

What is Statistics? A practical example

- Michael Benton & Francisco Ayala, *Dating the Tree of Life*, Science 2003 300: 1698-1700
- **Molecular** vs. **Paleontological** dating of major branching points in the tree of life are debated
- **Molecular** date estimates are up to twice as old (due to statistical bias) as **Paleontological** dates (missing fossils).
- **Goals**: Same as that set out by Darwin: to understand *where life came from*, the *shape of evolution*, the *place of humans in nature* and to determine the *extent of modern biodiversity* and where *it is threatened*.

What is Statistics? A practical example

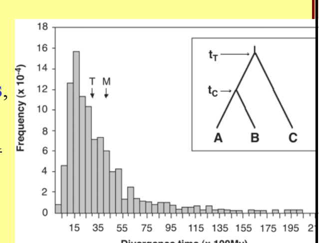
- **Plants**: The first vascular land plants are found as **fossils** in the Silurian, and earlier evidence from possible vascular plant spores may extend the range back to the Ordovician, **475 Ma** considerably < a molecular estimate of **700 Ma**.
- **Birds**: Molecular estimates place the split of basal clades and modern orders at 70 to 120 Ma. The oldest uncontroversial fossils of modern bird orders date from the Paleocene (60 Ma), much younger.
- **Mammals**: Molecular dates split of modern placentals in the mid- to Late Cretaceous (80 to 100 Ma). The oldest fossil representatives of modern mammals dated from the Paleocene and Eocene (50 to 65 Ma).

What is Statistics? Topics!

- It is proposed that molecular dates are correct (with **confidence intervals**) and that methods exist to correct for that **error**. However, critics have pointed out several **pervasive biases** that make molecular dates too old.
- First, if calibration dates are too old, then all other dates **estimated** from them will also be too old.
- A second biasing factor is that undetected fast-evolving genes could **bias** estimates of timing. **Empirical and statistical studies** of vertebrate sequences suggest that such non-clock-like genes may be detected and that they do not affect **estimates** of dating. However, **statistical tests** may have low **power** and could produce consistently > dates.

What is Statistics? A practical example

- A 3rd source of **bias** relates to polymorphism. Two species often become fixed for alternative alleles that existed as a polymorphism in their ancestral species.
- A 4th biasing **factor** is that molecular time estimates show (**skewed**) **asymmetric distributions**, with a constrained (large numbers) younger left-end and an unconstrained (smaller numbers) older right-end.



What is Statistics? Estimate Variation!

Data Source	Metazoa (Animals) In MYA	Bilateria (metazoans except sponges, e.g., anemones)	Deuterostomia (backboned animals)
Gene (8 G)		1200 ± 100	1001 ± 100
Protein (64 E)	930 ± 115	790 ± 60	590
Gene (4 G)	940 ± 80	700 ± 80	
Gene (18 G)		670 ± 60	600 ± 60
Gene (22 G)		830 ± 55	
Gene (50 G)	1350 ± 150 (est.)	993 ± 46	
Gene (22 G)		659 ± 131	
Protein (10 E)		627 ± 51	
Gene (MtDNA 18S rRNA)		588 min.	586/589 min.

molecular estimates are that basal splits among major animal clades happened about 1000 MYA

What is Statistics? A practical example

- *Demography: Uncertain population forecasts*

  by Nico Keilman, Nature 412, 490 - 491 (2001)
- Traditional population forecasts made by statistical agencies **do not quantify uncertainty**. But demographers and statisticians have developed methods to calculate probabilistic forecasts.
- The demographic future of any human population is uncertain, but some of the many possible trajectories are more probable than others. So, forecast demographics of a population, e.g., size by 2100, should include two elements: a range of possible outcomes, and a probability attached to that range.

What is Statistics?

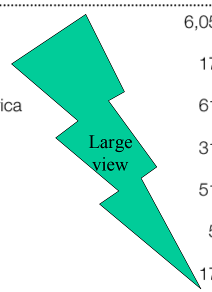
- Together, ranges/probabilities constitute a *prediction interval* for the population. There are trade-offs between greater certainty (higher odds) and better precision (narrower intervals). Why?
- For instance, the next table shows an estimate that the odds are 4 to 1 (an 80% chance) that the world's population, now at 6.1 billion, will be in the range [5.6 : 12.1] billion in the year 2100. Odds of 19 to 1 (a 95% chance) result in a **wider** interval: [4.3 : 14.4] billion.

Table 1 Forecasted population sizes and proportions over age 60

	Median world and regional population sizes (millions)				
Year	2000	2025	2050	2075	2100
World total	6,055	7,827 (7,219–8,459)	8,797 (7,347–10,443)	8,951 (6,636–11,652)	8,414 (5,577–12,123)
North Africa	173	257 (228–285)	311 (249–378)	336 (238–443)	333 (215–484)
Sub-Saharan Africa	611	976 (856–1,100)	1,319 (1,010–1,701)	1,522 (1,021–2,194)	1,500 (878–2,450)
North America	314	379 (351–410)	422 (358–498)	441 (343–565)	454 (313–631)
Latin America	515	709 (643–775)	840 (679–1,005)	904 (647–1,202)	934 (585–1,383)
Central Asia	56	81 (73–90)	100 (80–121)	107 (76–145)	106 (66–159)
Middle East	172	285 (252–318)	368 (301–445)	413 (296–544)	413 (259–597)
South Asia	1,367	1,940 (1,735–2,154)	2,249 (1,795–2,776)	2,242 (1,528–3,085)	1,958 (1,186–3,035)
China region	1,408	1,608 (1,494–1,714)	1,580 (1,305–1,849)	1,422 (1,003–1,884)	1,250 (765–1,870)
Pacific Asia	476	625 (569–682)	702 (575–842)	702 (509–937)	654 (410–949)
Pacific OECD	150	155 (144–165)	148 (125–174)	135 (100–175)	123 (79–173)
Western Europe	456	478 (445–508)	470 (399–549)	433 (321–562)	392 (257–568)
Eastern Europe	121	117 (109–125)	104 (86–124)	87 (61–118)	74 (44–115)
European part of the former USSR	236	218 (203–234)	187 (154–225)	159 (110–216)	141 (85–218)

Table 1 Forecasted population sizes and proportions over

	Median	
Year	2000	2025
World total	6,055	7,827 (7,219–8,459)
North Africa	173	257 (228–285)
Sub-Saharan Africa	611	976 (856–1,100)
North America	314	379 (351–410)
Latin America	515	709 (643–775)
Central Asia	56	81 (73–90)
Middle East	172	285 (252–318)



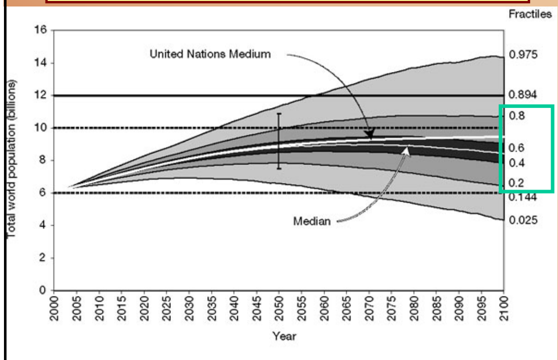
What is Statistics?

- *Demography: Uncertain population forecasts*

  by Nico Keilman, Nature 412, ,2001
- Traditional population forecasts made by statistical agencies **do not quantify uncertainty**. But lately demographers and statisticians have developed methods to calculate probabilistic forecasts.
- Proportion of population over 60yrs.

Proportion of population over age 60		
2000	2050	2100
0.10	0.22 (0.18–0.27)	0.34 (0.25–0.44)
0.06	0.19 (0.15–0.25)	0.32 (0.23–0.44)
0.05	0.07 (0.05–0.09)	0.20 (0.14–0.27)
0.16	0.30 (0.23–0.37)	0.40 (0.28–0.52)
0.08	0.22 (0.17–0.28)	0.33 (0.23–0.45)
0.08	0.20 (0.15–0.25)	0.34 (0.24–0.46)
0.06	0.18 (0.14–0.23)	0.35 (0.24–0.47)
0.07	0.18 (0.14–0.24)	0.35 (0.25–0.48)
0.10	0.30 (0.24–0.37)	0.39 (0.27–0.53)
0.08	0.23 (0.18–0.29)	0.36 (0.26–0.49)
0.22	0.39 (0.32–0.47)	0.49 (0.35–0.61)
0.20	0.35 (0.29–0.43)	0.45 (0.32–0.58)
0.18	0.38 (0.30–0.46)	0.42 (0.28–0.57)
0.19	0.35 (0.27–0.44)	0.36 (0.23–0.50)

What is Statistics?



What is Statistics?

- There is concern about the accuracy of population forecasts, in part because the rapid fall in fertility in Western countries in the 1970s came as a surprise. Forecasts made in those years predicted birth rates that were up to 80% too high.
- The rapid reduction in mortality after the Second World War was also not foreseen; life-expectancy forecasts were too low by 1–2 years; and the predicted number of elderly, particularly the oldest people, was far too low.

What is Statistics?

- So, during the 1990s, researchers developed methods for making probabilistic population forecasts, the **aim** of which is to calculate prediction intervals for every variable of interest. Examples include population forecasts for the USA, AU, DE, FIN and the Netherlands; these forecasts comprised prediction intervals for variables such as age structure, average number of children per woman, immigration flow, disease epidemics.
- We need accurate probabilistic population forecasts for the whole world, and its 13 large division regions (see Table). The conclusion is that there is an estimated 85% chance that the world's population will stop growing before 2100. Accurate?

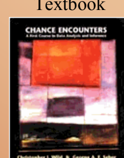
What is Statistics?

- There are three main methods of probabilistic forecasting:
  - time-series extrapolation;
  - expert judgement; and
  - extrapolation of historical forecast errors.
- Time-series methods rely on statistical models that are fitted to historical data. These methods, however, seldom give an accurate description of the past. If many of the historical facts remain unexplained, time-series methods result in excessively wide prediction intervals when used for long-term forecasting.
- Expert judgement is subjective, and historic-extrapolation alone may be near-sighted.

Chapter 1: What is Statistics?

Chris Wild & George Seber

Textbook



- Polls and surveys – we're all different; It's impossible or expensive to investigate every single person.
- Experimentation – sample vs. population
- Observational Studies – selection and non-response bias
- Statistics -- What is it and who uses it?
- Summary

Newtonial science vs. chaotic science

- Article by Robert May, Nature, vol. 411, June 21, 2001
  - Science we encounter at schools deals with crisp certainties (e.g., prediction of planetary orbits, the periodic table as a descriptor of all elements, equations describing area, volume, velocity, position, etc.)
  - As soon as uncertainty comes in the picture it shakes the foundation of the deterministic science, because only probabilistic statements can be made in describing a phenomenon (e.g., roulette wheels, chaotic dynamic weather predictions, Geiger counter, earthquakes, etc.)
  - What is then science all about – describing absolutely certain events and laws alone, or describing more general phenomena in terms of their behavior and chance of occurring? Or may be both!

Variation in sample percentages

Poll: Do you consider yourself overweight?

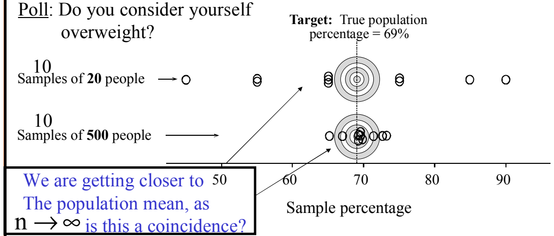


Figure 1.1.1 Comparing percentages from 10 different surveys each of 20 people with those from 10 surveys each of 500 people (all surveys from same population).

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Errors in Samples …

- Selection bias: Sampled population is not a representative subgroup of the population really investigated.
- Non-response bias: If a particular subgroup of the population studied does not respond, the resulting responses may be skewed.
- Question effects: Survey questions may be slanted or loaded to influence the result of the sampling.
- Is quota sampling reliable? Each interviewer is assigned a fixed quota of subjects (subjects district, sex, age, income exactly specified, so investigator can select those people as they liked).
- *Target population* –entire group of individuals, objects, units we study.
- *Study population* –a subset of the target population containing all "units" which could possibly be used in the study.
- *Sampling protocol – procedure used to select* ***the sample***
- *Sample –* the subset of "units" about which we actually collect info.

More terminology …

- ***Census*** *– attempt to sample the entire population*
- ***Parameter*** *– numerical characteristic of the* ***population****, e.g., income, age, etc. Often we want to* ***estimate population parameters****.*
- ***Statistic*** *– a numerical characteristic of the* ***sample. (Sample) statistic*** ***is used to estimate a corresponding population parameter.***
- Why do we sample at random? We draw "units" from the study population at random to avoid bias. Every subject in the study sample is equally likely to be selected. Also random-sampling allows us to calculate the likely size of the error in our sample estimates.

More definitions …

- How could you implement the lottery method to randomly sample 10 students from a class of 250? – list all names; assign numbers 1,2,3,…,250 to all students; Use a random-number generator to choose (10-times) a number in range [0;250]; Process students drawn.
- Random or chance error is the difference between the sample-value and the true population-value (e.g., 49% vs. 69%, in the above body-overweight example).
- Non-sampling errors (e.g., non-response bias) in the census may be considerably larger than in a comparable survey, since surveys are much smaller operations and easier to control.
- Sampling errors–arising from a decision to use a sample rather than entire population
- ***Unbiased procedure/protocol:*** (e.g., using the proportion of left-handers from a random sample to estimate the corresponding proportion in the population).
- **Cluster sampling**- a cluster of individuals/units are used as a sampling unit, rather than individuals.

More terminology …

- What are some of the ***non-sampling errors*** that plague surveys? (non-response bias, question effects, survey format effects, interviewer effects)
- If we take a random sample from one population, can we apply the results of our survey to other populations? (It depends on how similar, in the respect studied, the two populations are. In general- No! This can be a dangerous trend.)
- Are sampling households at random and interviewing people at random on the street valid ways of sampling people from an urban population? (No, since clusters (households) may not be urban in their majority.)
- Pilot surveys – after prelim investigations and designing the trial survey Q's, we need to get a "small sample" checking clearness and ambiguity of the questions, and avoid possible sampling errors (e.g., bias).

Questions …

- How do the following lead to biases or cause differences in response:
  - ***non-response***
  - ***self-selection***
  - ***question effects***
  - ***survey-format effects***
  - ***interviewer effects***
  - ***transferring findings***?

Questions …

- Give an example where non-representative information from a survey may be useful. Non-representative info from surveys may be used to estimate parameters of the actual sub-population which is represented by the sample. E.g., Only about 2% of dissatisfied customers complain (most just avoid using the services), these are the most-vocal reps. So, we can not make valid conclusions about the stereotype of the dissatisfied customer, but we can use this info to tract down changes in levels of complains over years.
- Why is it important to take a pilot survey?
- Give an example of an unsatisfactory question in a questionnaire. (In a telephone study: What time is it?

Do we mean Eastern/Central/Mountain/Pacific?)

Questions …

- Random allocation – randomly assigning treatments to units, leads to representative sample only if we have large # experimental units.
- Completely randomized design- the simplest experimental design, allows comparisons that are unbiased (not necessarily fair). Randomly allocate treatments to all experimental units, so that every treatment is applied to the same number of units. E.g., If we have 12 units and 3 treatments, and we study treatment efficacy, we randomly assign each of the 3 treatments to 4 units exactly.
- Blocking- grouping units into blocks of similar units for making treatment-effect comparisons only within individual groups. E.g., Study of human life expectancy perhaps income is clearly a factor, we can have high- and low-income blocks and compare, say, gender differences within these blocks separately.

Questions …

- Why should we try to "blind" the investigator in an experiment?
- Why should we try to "blind" human experimental subjects?
- The basic rule of experimentor :

"Block *what you can and* *randomize* *what you cannot.*"

Experiments vs. observational studies for comparing the effects of treatments

- In an Experiment
  - experimenter determines which units receive which treatments. (ideally using some form of random allocation)
- Observational study – useful when can't design a controlled randomized study
  - compare units that happen to have received each of the treatments
  - Ideal for describing relationships between different characteristics in a population.
  - often useful for identifying possible causes of effects, but cannot reliably establish causation.
- Only properly designed and executed experiments can reliably demonstrate causation.

Questions …

- What is the difference between a designed experiment and an observational study? (no control of the design in observational studies)
- Can you conclude causation from an observational study? Why or why not? (not in general!)
- How do we try to investigate causation questions using observational studies? In a smoking-lung-cancer study: try to divide all subjects, in the obs. study, into groups with equal, or very similar levels of all other factors (age, stress, income, etc.) – I.e. control for all outside factors. If rate of lung-cancer is still still higher in smokers we get a stronger evidence of causality.
- What is the idea of controlling for a variable, and why is it used? Effects of this variable in the treatment/control groups are similar.
- Epidemiology – science of using statistical methods to find causes or risk factors for diseases.

The Subject of Statistics

Statistics is concerned with the process of finding out about the world and how it operates -

- in the face of variation and uncertainty
- by collecting and then making sense (interpreting) of data.

The investigative process

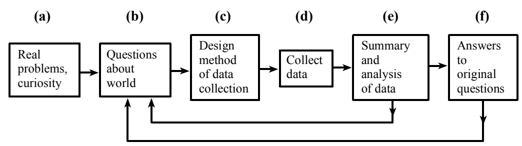


Figure 1.4.1 The investigative process.

CHAPTER SUMMARY

Experiments and observational studies

- When exploring questions of cause and effect we distinguish between ***observational studies*** and ***experiments***.
  - In an ***experiment***, the experimenter determines which subjects (experimental units) receive which treatments
  - In an ***observational study***, we simply compare subjects that happen to have received each of the treatments.
  - Observational studies widely used for identifying possible causes of effects but cannot reliably establish causation
  - Only properly designed and executed experiments (Section 1.2) can reliably demonstrate causation.

The Role of Randomization

Well designed statistical studies employ randomization to avoid subjective and other biases.

- Surveys and observational studies should use random sampling to obtain representative samples.
- Experiments should use random assignment of experimental subjects to treatment groups
  - to ensure comparisons are fair i.e., treatment groups are as similar as possible in every way except for the treatment being used.

"Blocking" vs. "stratification"

"***Blocking***"

- word used in describing an experimental design

"***Stratification***"

- used in describing a survey or observational study

- Both refer to idea of only making comparisons within relatively similar groups of subjects

Blocking and randomization

"***Block what you can and randomize what you cannot***."

- Block to ensure fair comparisons with respect to factors known to be important
- Randomize to try to obtain comparability with respect to unknown factors
- Randomization also allows the calculation of how much the estimates made from the study data are likely to be in error

Sources of error in surveys

- Random sampling leads to ***sampling errors***, sampling-size (as we saw for the overweight survey), arising for the choice to use a sample, as opposed to census.
- ***Non-sampling errors*** can be much larger than the sampling errors. Selection bias, non-response bias, survey/question/interview format are all non-sampling errors.

Sources of non-sampling errors

- ***Selection bias***:
  Arises when the population sampled is not exactly the population of interest.
- ***Self-selection***:
  People themselves decide whether or not to be surveyed. Results akin to severe non-response.
- ***Non-response bias***:
  Non-respondents often behave or think differently from respondents
  - low response rates can lead to huge biases.

Non-sampling errors cont.

- ***Question-wording effects***:
  Even slight differences in question wording can produce measurable differences in how people respond.
- ***Interviewer effects***:
  Different interviewers asking the same questions can tend to obtain different answers.
- ***Survey format effects***:
  Factors such as question order, questionnaire layout, self-administered questionnaire or interviewer, can effect the results.

Dealing with errors

- ***Statistical methods*** are available for estimating the likely size of ***sampling errors***.
- All we can do with ***non-sampling errors*** is to try to minimize them at the study-design stage.
- ***Pilot survey***:
  One tests a survey on a relatively small group of people to try to identify any problems with the survey design before conducting the survey proper.

Jargon describing experiments

- ***Control group***:
  - group of experimental units is given no treatment.
  - treatment effect estimated by comparing each treatment group with control group
- ***Blinding***:
  - Preventing people involved in experiment from knowing which experimental subjects have received which treatment
  - One may be able to blind
    - subjects themselves
    - people administering the treatments
    - people measuring the results.

Jargon describing experiments

- ***Double blind***:
  Both the subjects and those administering the treatments have been blinded.
- ***Placebo***:
  An inert/dummy/fake treatment.
- ***Placebo effect***:
  Response caused in human subjects by the idea that they are being treated.

Immigration Example

- Suppose that you want to set up a nationwide survey about immigration issues. Think as precisely as you can about the target population that you would be interested in.

  -Who would you want included?

  -Who would you want excluded?

  -Can you define some rules to characterize your target population?

Immigration Example

- We could take all members of the population in the US at the time, who were entitled to vote in national elections. This may exclude the young, the illegal immigrants, those people in prisons and people legally committed to mental institutions. It would include any other permanent residents of the US, whether or not they were citizens, and citizens living overseas.
- You might want to be more, or less, restrictive. In practice, one would probably sample from something like the electoral
- districts [that subset of people who fit the eligibility criteria for voting and who have registered to do so].
- Should the goals of the study influence your survey design (in particular how conservative your selection is)?

Poll Example

- A survey of High School principals taken after a widespread change in the public school system revealed that 20% of them were under stress-reliefe medication, and almost 50% had seen a doctor in the past 6 mo.s with stress complains. The survey was compiled from 250 questionnaires returned out of 2500 sent out. How reliable the results of this experiment are and why?

Poll Example

- This is only a 10% response rate - the people who responded could be very unrepresentative. It could well be that the survey struck a responsive chord with stressed-out principals.

Experimental vs. Observation study

- A researcher wants to evaluate IQ levels are related to person's height. 100 people are are randomly selected and grouped into 5 bins: [0:50), [50;100), [100:150], [150:200), [200:250] *cm* in height. The subjects undertook a IQ exam and the results are analyzed.
- Another researcher wants to assess the bleaching effects of 10 laundry detergents on 3 different colors (R,G,B). The laundry detergents are randomly selected and applied to 10 pieces of cloth. The discoloration is finally evaluated.

Experimental vs. Observation study

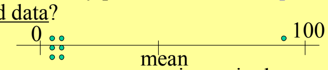
- For each study, describe what *treatment* is being compared and what *response* is being measured to compare the treatments.
- Which of the studies would be described as *experiments* and which would be described as *observational* studies?
- For the studies that are observational, could an experiment have been carried out instead? If not, briefly explain why not.
- For the studies that are experiments, briefly discuss what *forms of blinding* would be possible to be used.
- In which of the studies has *blocking* been used? Briefly describe *what* was blocked and why it was blocked.

Experimental vs. Observation study

- What is the *treatment* and what is the *response?*
  1. Treatment is height (as a bin). Response is IQ score.
  2. Treatment is laundry detergent. Response is discoloration.
- *Experiment* or *observational* study?
  1. Observational – compare obs's (IQ) which happen to have the treatment (height).
  2. Experimental – experimenter controls which treatment is applied to which unit.
- For the observational studies, can we conduct an experiment?
  1. This could not be done as an experiment - it would require the experimenter to decide the (natural) height (treatment) of the subjects (units).
- For the experiments, is there *blinding?*
  2. The only form of blinding possible would be for the technicians measuring the cloth discoloration not to know which detergent was applied.
- Is there *blocking*?
  1. & 2. No blocking. Say, if there are two laundry machines with different cycles of operation and if we want to block we'll need to randomize which laundry does which cloth/detergent combinations, because differences in laundry cycles are a known source of variation

Mean, Median, Mode, Quartiles, 5# summary

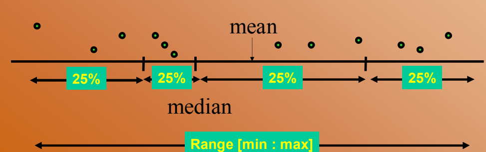
- The sample mean is the average of all numeric obs's.
- The sample median is the obs. at the index $(n+1)/2$ (note take avg of the 2 obs's in the middle for fractions like 23.5), of the observations ordered by size (small-to-large)?
- The sample median usually preferred to the sample mean for skewed data?



- Under what circumstances may quoting a single center (be it mean or median) not make sense?(multi-modal)
- What can we say about the sample mean of a qualitative variable? (meaningless)

Quartiles

The first quartile (Q_1) is the median of all the observations whose *position* is strictly below the *position* of the median, and the third quartile (Q_3) is the median of those above.



Five number summary

The five-number summery = (Min, Q_1, Med, Q_3, Max)