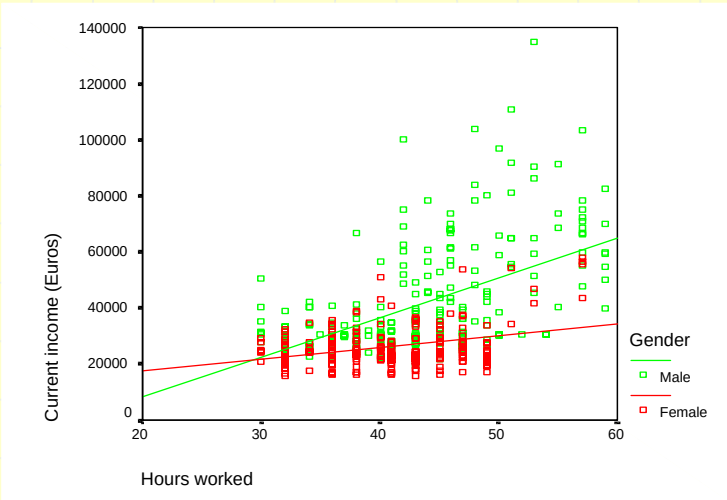
UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH
STATISTICAL CONSULTING UNIT

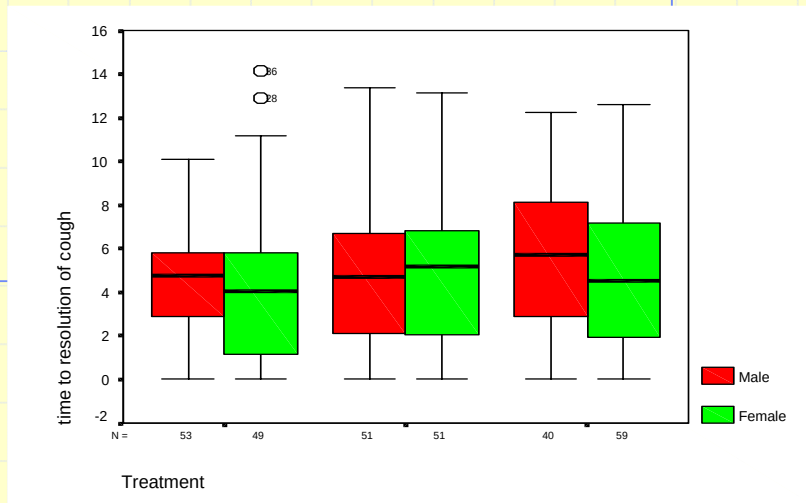
Research in Health Sciences

Medical Statistics

Dr Jean Saunders C.Stat

Statistical Consulting Unit (SCU)





Outline of talk

- Probability and Statistics
- Data summarisation and Inference
- Sample Size calculations (Reminder)
- Statistical Consulting Unit (SCU)
- Contact details

Probability/Uncertainty

Source of Uncertainty :- **Variability**

Without **variability** events will be entirely predictable

1. Survival time after a heart transplant will vary from patient to patient
2. Response to a new drug will vary from patient to patient
3. Not all smokers will develop lung cancer

Statistical methods assign a degree of uncertainty to those answers

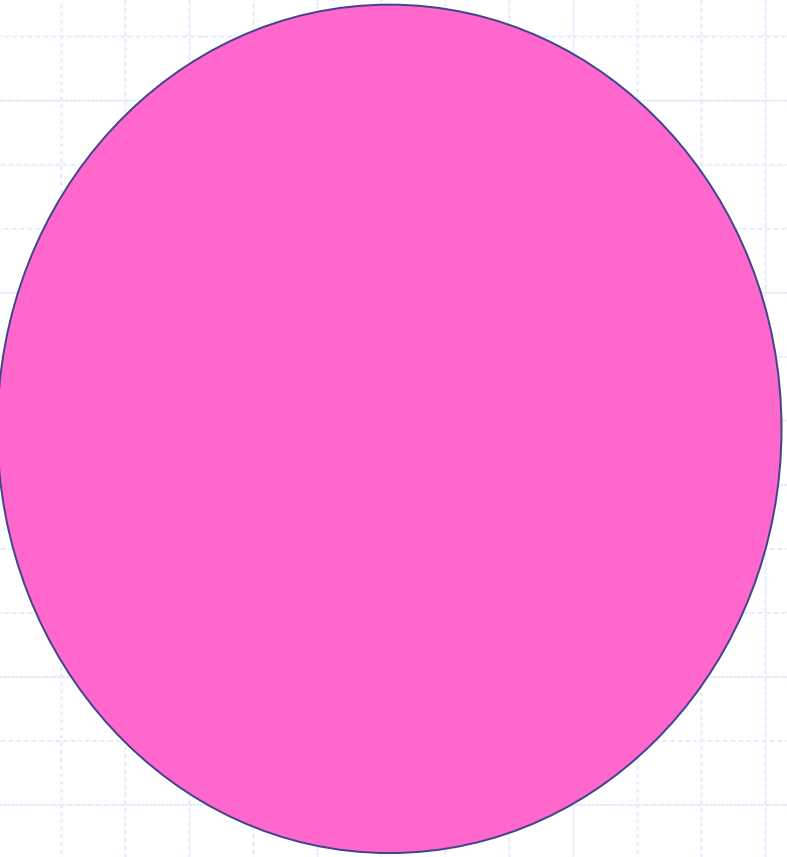
Example (more details later)

24 hour total energy expenditure(MJ/day) in groups of lean and obese women (Prentice et al., 1986)

Lean (n=13)	6.13	7.05	7.48	7.48	7.53	7.58
7.90	8.08	8.09	8.11	8.40	10.15	10.88
Obese(n=9)	8.79	9.19	9.21	9.68	9.69	9.97
11.51	11.85	12.79				

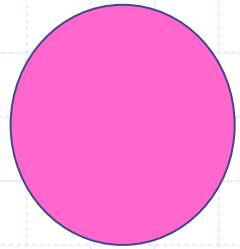
p-value = 0.0008

If the energy expenditure in Lean and Obese women is really the same, then the probability of obtaining the difference as shown by the above data (or a larger difference) is 0.0008 (or 8 in 10000).



From population to sample

From sample to population



Sample

Population

When a sample is drawn there is no certainty that a particular individual will be selected from the population

1. Individuals in the population vary from one another with respect to an outcome of interest

Examples:

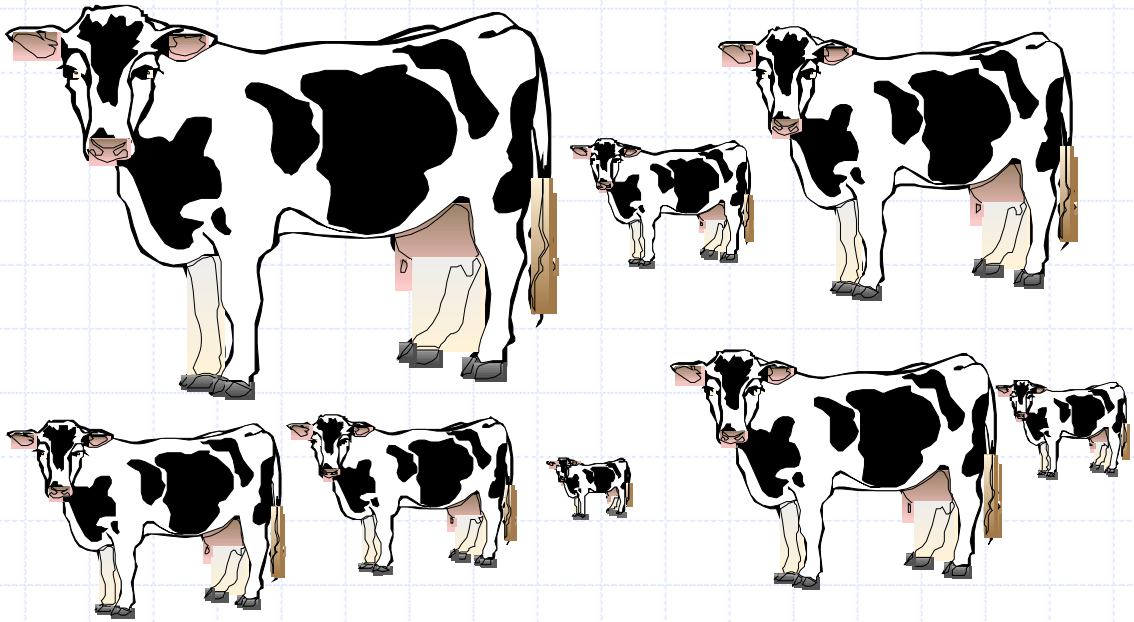
(a) Survival time after heart transplant will vary from patient to patient

(b) Response to a drug will vary from patient to patient.

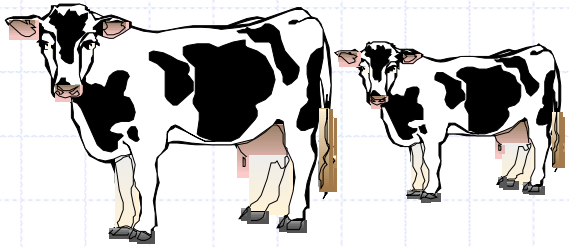
2. When a sample is drawn there is no certainty that a particular individual will be selected from the population

3. Thus, the samples also vary from each other in representation of the population from which they are drawn

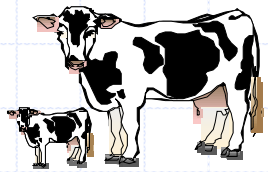
Generally the population is not homogeneous



Giving rise to variability in samples

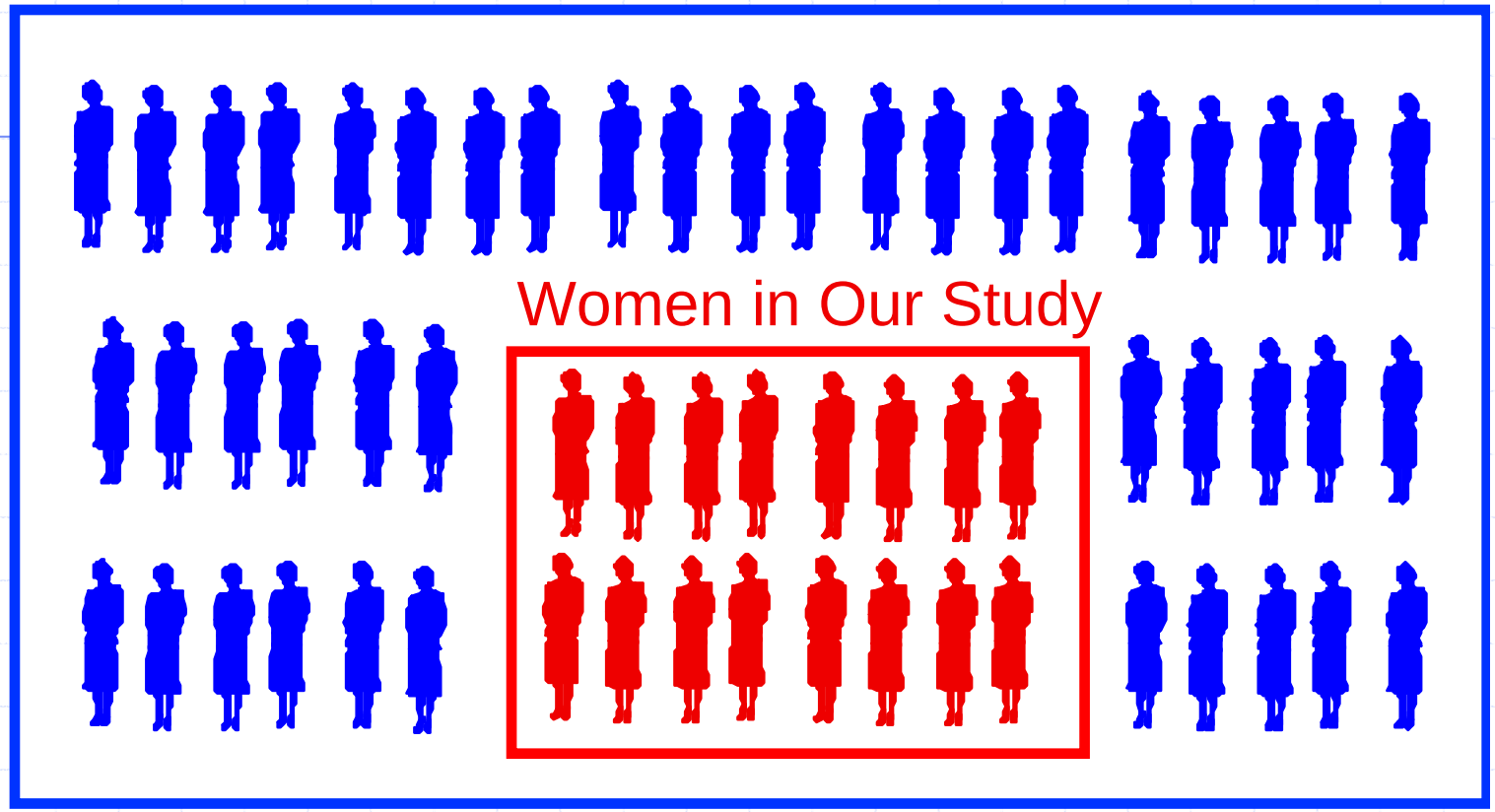


Sample 1



Sample 2

All Women with Breast Cancer



Population: A class of patients about whom we would like to draw conclusions.

Sample: A group of patients from the population whose outcome is known.

Probability at work

1. A sample is drawn from a population to make inference about the population (say, to see if the cure rate is more than 50%)

2. This is one of the several possible samples from the population. Some samples resemble the population closely; some resemble it poorly.

3. For a large sample size, probability of obtaining a sample that closely resembles the population characteristic is higher. Probability of obtaining a sample with poor resemblance of the population is small.

Basic Ideas Behind Hypothesis Testing

We take a sample from the population and calculate statistics assuming that our sample comes from a particular population

(Null hypothesis)

Decision:

If the sample resembles the population poorly, then we conclude that the sample did not come from the population of our interest, but it came from some other population.

(Generally a probability of 0.05 or less is considered small enough. This is also called the level of significance)

Note: Air travel is not absolutely safe, but the chance of accidents are so small that we consider it quite safe. The accident data are unlikely to come from an 'unsafe situation'.

- **Data Summarisation**
  Descriptive Statistics

  How do we summarise complex results of clinical studies in a meaningful way?

- **Nature of Statistical Inference**
  Hypothesis Testing

  We give a new therapy to 100 breast cancer patients. What can we conclude from this study about how other breast cancer patients should be treated? Is it better than the old therapy?

Types of Data

Categorical, Numerical (qualitative, quantitative)

***Categorical data*:**

Observation on an individual in the sample is made by assigning the individual to one of several categories or classifications

Example 1: Blood type - A, B, O, AB
Example 2: Cancer Stage - I, II, III, IV
Example 3: Gender - Female, Male
Example 4: Smoking - Yes, No
Example 5: Smoking - No, Light, Medium, Heavy

Examples 1, 3 and 4 are examples of *nominal categorical data*
Examples 2 and 5 are examples of *ordered categorical data*

***Numerical Data*:**

Generally, observations are made by measuring or by counting

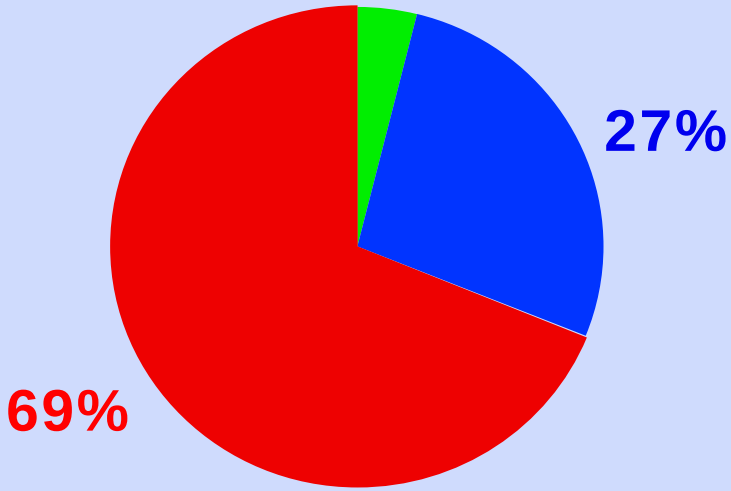
Example 1: Number of live births
Example 2: Weight, Height, Blood pressure

Example 1 is an example of discrete numerical data.

Example 2 is an example of continuous numerical data.

Methods of Describing Data

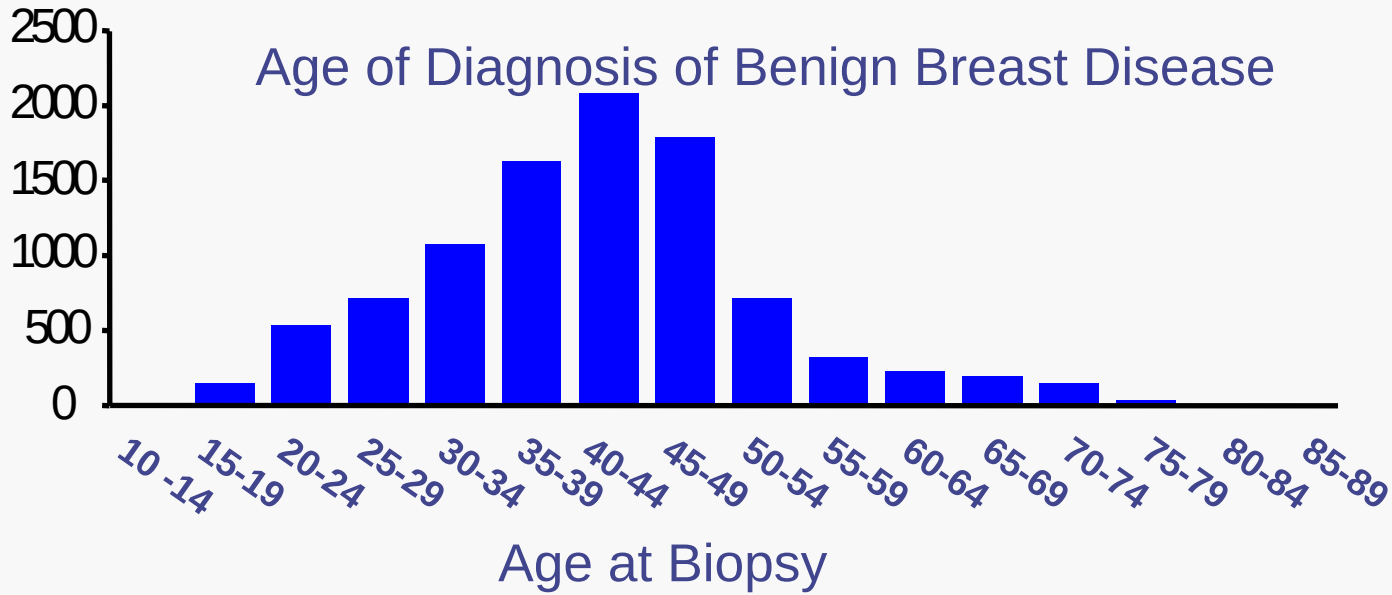
- Pie Charts
- Bar Charts
- Histogram

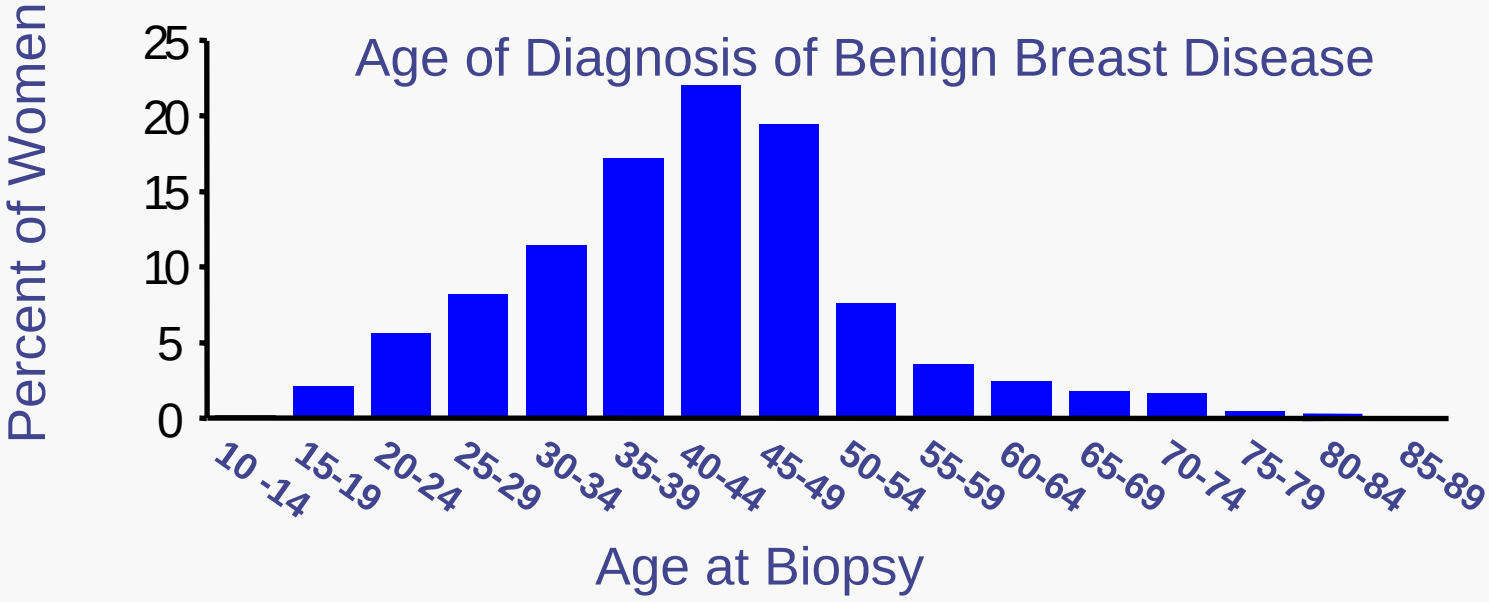


Atypical Hyperplasia

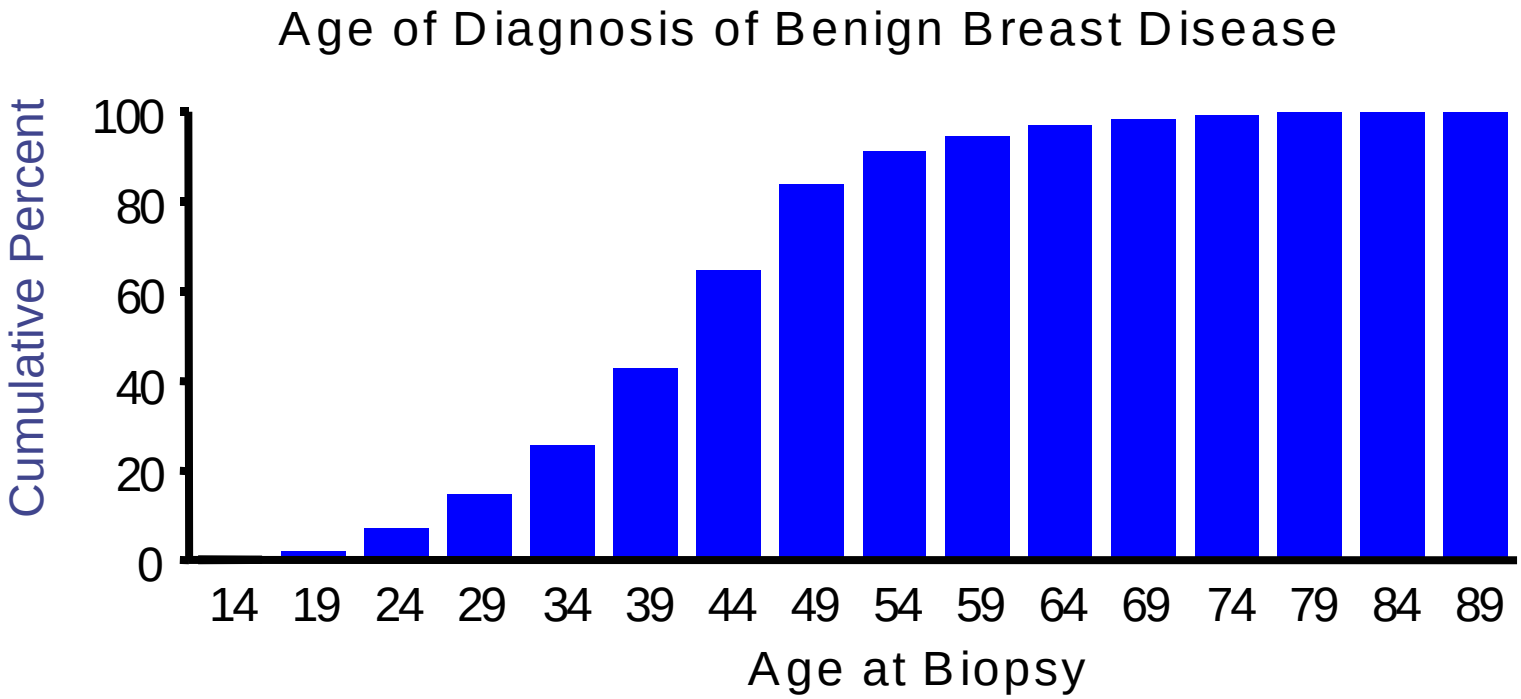
Proliferative Disease Without Atypia

No Proliferative Disease





Mode: Most common value observed



Cumulative Frequency: The cumulative frequency by age *K* is the proportion of patients whose age of biopsy is $\leq K$.

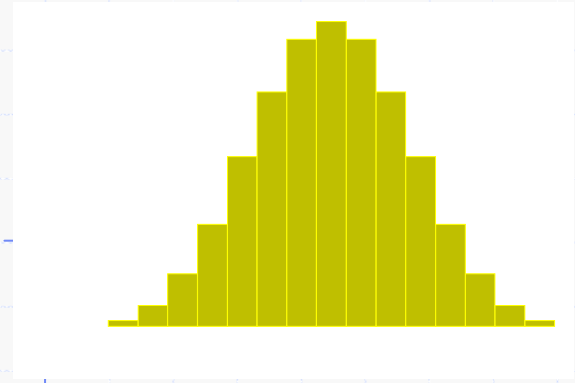
E.G. 43% of patients were biopsied by age 39. The cumulative frequency by age 39 is 43%.

Percentile: The p-th percentile is that age whose cumulative frequency is p.

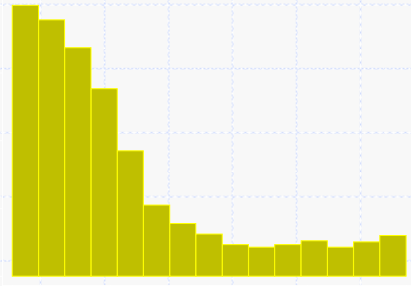
E.G. The 43rd percentile of the age of biopsy is 39.

Median: The 50th percentile.

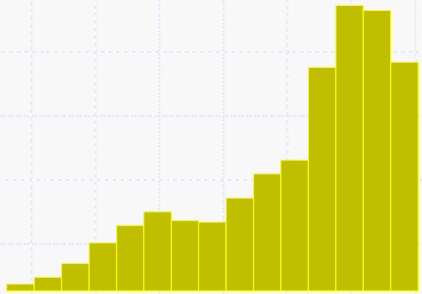
Histograms



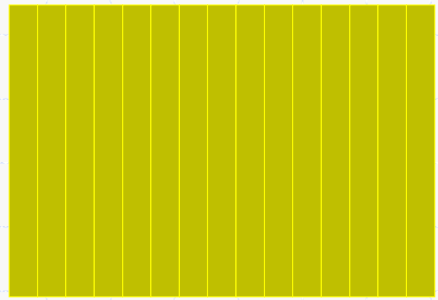
Bell -shaped



Skewed to the right



Skewed to the left



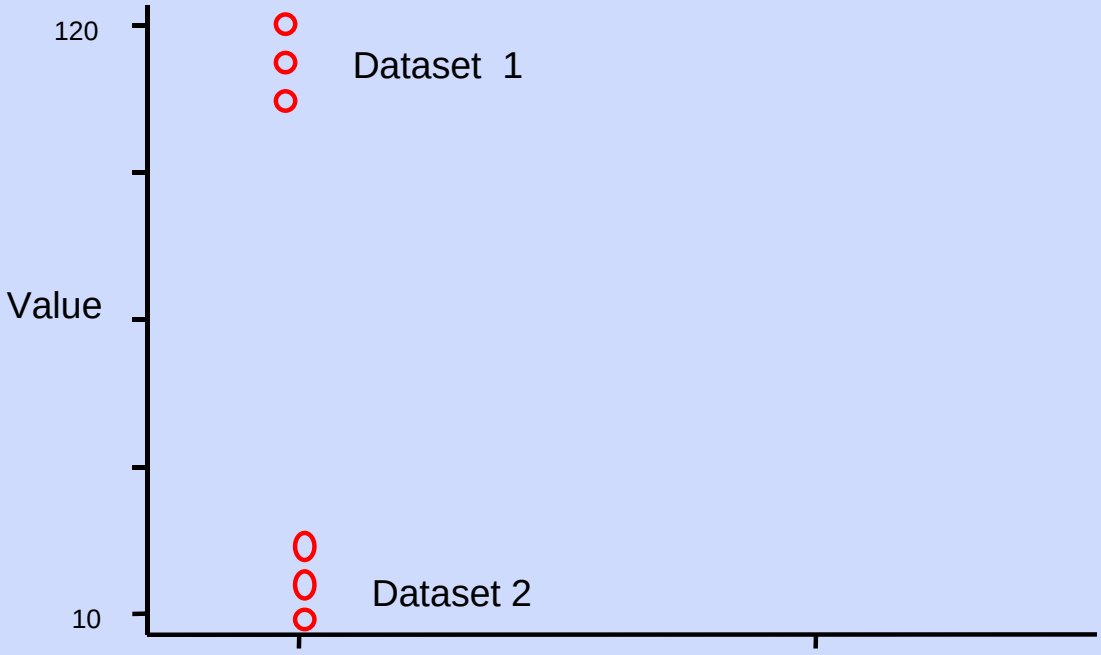
Uniform

A line through to mid-points of bars of a bell shaped histogram

Numerical summarisation

Measures of central tendency or (measures of location)

Data set 1	Data set 2
110	20
115	10
120	15

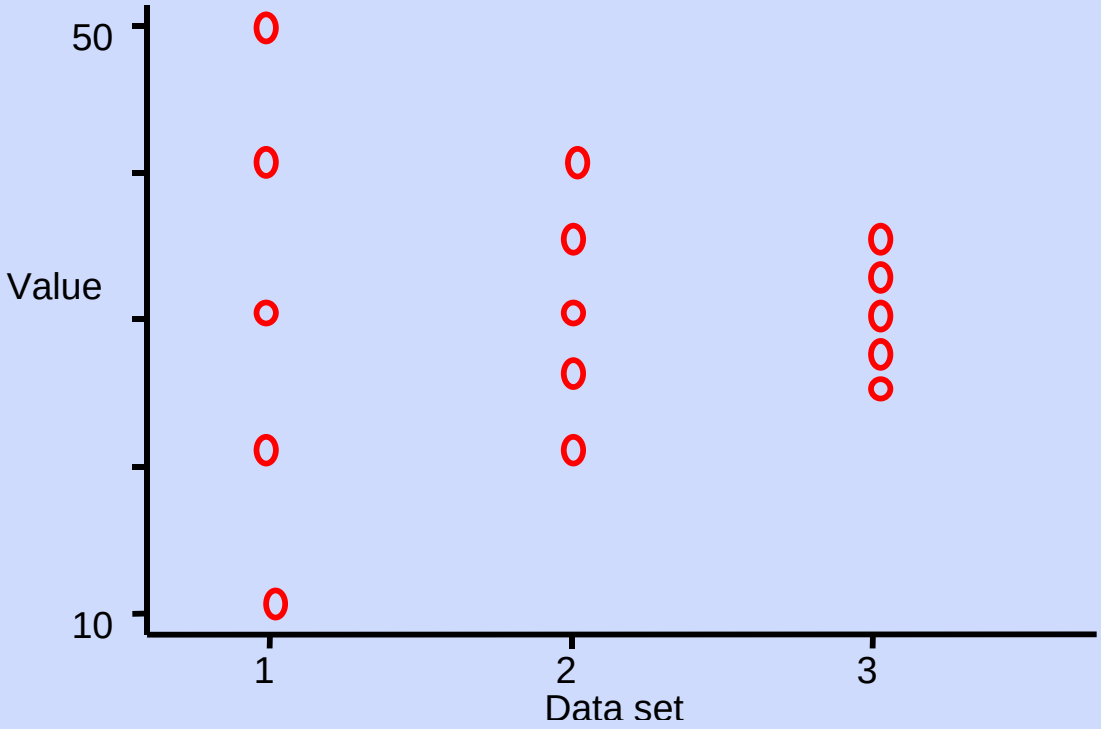


Measures of Central Tendency

Mean: Average Value
Median: 50th percentile
Mode: Most common value

Mean can be overly affected by extreme values.

Data set	Observations	Mean	Median
1	10, 20, 30, 40, 50	30	30
2	20, 25, 30, 35, 40	30	30
3	28, 29, 30, 31, 32	30	30



Measures of Variability

Range: Largest - smallest observed value.

- The range increases with increasing sample size.
- Overly affected by extreme values

Interquartile Range: 75th percentile - 25th percentile

Semi-Interquartile Range: IQR/2

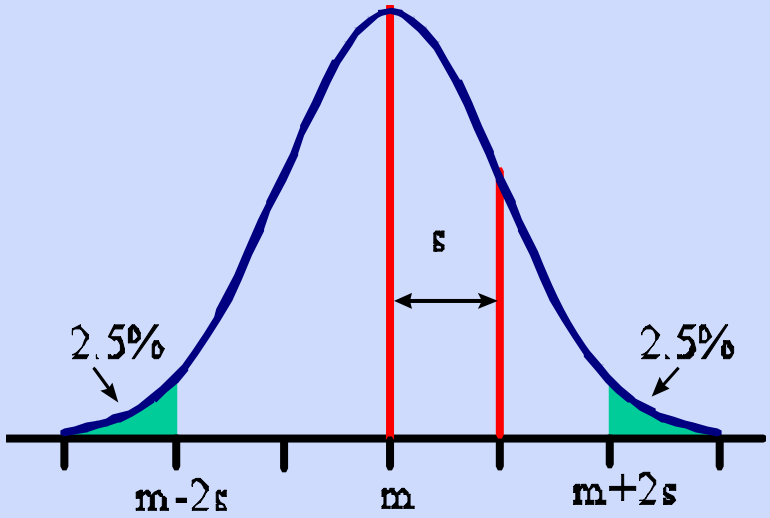
(Sample) Standard Deviation (s):

- Badly affected by extreme values
- A mathematician friendly statistic

Mean and Standard Deviation Together

For a normal distribution

- 68 % of observations lie within one s.d. of the mean
- 95 % of observations lie within two s.d. of the mean
- 99.7 % of observations lie within three s.d. of the mean



Probability of a sample being drawn from a population may be obtained from this curve.

Inference

Point estimation

-Using a single value to estimate the population parameter

Interval estimation

-Using an interval to estimate the population parameter

Hypothesis testing

-Making an hypothesis about the population parameter and see if the sample disproves it.

Point Estimation

Sample mean is a point estimate of the population mean.

Interval Estimation

A 95% Confidence Interval for a parameter will contain the true value of the parameter 95% of the time.

- The 95% confidence interval for a population mean obtained from a sample of size N

  = Mean ± 2 SE

  = Mean ± 2 SD / $\sqrt{N}$

- In calculating confidence intervals we estimate the true standard deviation from the sample.

E.G. Sample mean = 45

Sample standard deviation = 30

Sample size N = 100

95% confidence interval = 45 ± 2 x 30 / $\sqrt{100}$

= 45 ± 2 X 3

= (39 – 51)

Hypothesis testing

Example 1: MI and OC Use

The following data (Rosener, 1995, table 10.2) classifies women 40-44 years of age by oral contraceptive (OC) use and incidence of myocardial infarction (MI) over a 3-year period.

	MI Status		
OC Use	*Yes*	*No*	Total
Yes	13	4987	5000
No	7	9993	10,000
Total	20	14,980	15,000

In this sample, risk of MI in users is about 3.71 times the risk of MI in non-user. $(13/5000)/(7/10000) = 26/7 = 3.71$

OC * MI Crosstabulation

			MI		
			yes	no	Total
OC	yes	Count	13	4987	5000
		Expected Count	6.7	4993.3	5000.0
		% within MI	65.0%	33.3%	33.3%
	no	Count	7	9993	10000
		Expected Count	13.3	9986.7	10000.0
		% within MI	35.0%	66.7%	66.7%
Total		Count	20	14980	15000
		Expected Count	20.0	14980.0	15000.0
		% within MI	100.0%	100.0%	100.0%

If we add in the column percentages to this table and look at the number expected in each cell if there was NO association (i.e. the null hypothesis) we can see that in fact there were twice as many who had experienced MI using OC's than we would expect with NO association.

Hypothesis testing

Example 1 (cont) : MI and OC Use

Null hypothesis: Risk of MI is the same in two groups

Alternate hypothesis: The risk is different in the two groups

Observed data (from a sample)

	MI Status		
OC Use	*Yes*	*No*	Total
Yes	13	4987	5000
No	7	9993	10,000
Total	20	14,980	15,000

P-value = 0.006 using the χ^2 test of association

Conclusion

Reject the null hypothesis

Chance of obtaining a sample showing at least this magnitude of relative risk is small (0.006). So we conclude that there is a higher risk with OC use. (BUT remember the data could have come from a population where there was no association with probability 0.006! Or there may be other factors involved)

Relative Risk

$$RR = \frac{\text{Risk in treatment group}}{\text{Risk in control group}}$$

$$= \frac{\text{Risk of MI in OC group}}{\text{Risk of MI in non-OC group}}$$

$$= \frac{13/5000}{7/10000}$$

$$= 3.71$$

Odds Ratios

(similar idea to odds in betting)

$$OR = \frac{\text{Odds in treatment Group}}{\text{Odds in Control Group}}$$

$$= \frac{\text{Odds of MI in OC group}}{\text{Odds of MI in non-OC group}}$$

$$= \frac{13/4987}{7/9993}$$

$$= 3.72$$

For rare outcomes (as in this case) RR and OR almost identical

Comparison of Relative Risk and Odds Ratios

Values of the RR when OR=2 and OR when RR=2 for different values of risk in the unexposed group

OR=2		RR=2	
Risk in Unexposed group	Corresponding RR	Risk in Unexposed group	Corresponding OR
0.001	1.998	0.001	2.002
0.01	1.980	0.01	2.020
0.1	1.818	0.1	2.25
0.9	1.053	0.3	6.00
0.99	1.005	0.5	∞

Example 2

Time spent in bed by 14 years old boys and girls

Null Hypothesis: Time spent in bed is independent of gender

Alternate hypothesis: Time spent in bed is not independent of gender

Time spent in bed (hours)

	7.0	7.5	8.0	8.5	9.0	9.5	10	>10	Total
Boys (obs)	88	109	210	324	359	313	182	85	1670
(exp)	*(87)*	*(104)*	*(206)*	*(324)*	*(383)*	*(311)*	*(183)*	*(72)*	
Girls (obs)	92	108	217	349	436	334	198	65	1799
(exp)	*(93)*	*(113)*	*(221)*	*(349)*	*(412)*	*(336)*	*(197)*	*(78)*	
Total	180	217	427	673	795	647	380	150	3469

p-value = 0.35 (using the χ^2 test of association)

Therefore we could not reject the null hypothesis.

BUT is this because there was really no difference between boys and girls or because the the test did not have enough power to detect the difference ?

(Is there a trend?)

Power: Probability of obtaining a statistically significant result if this alternative hypothesis is true.

Example 3

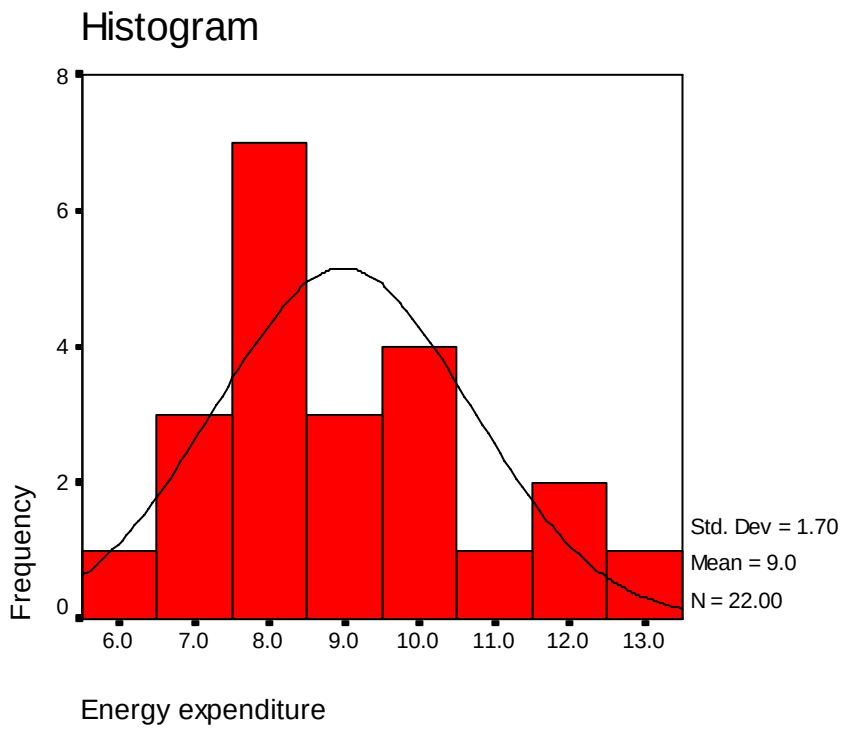
24 hour total energy expenditure(MJ/day) in groups of lean and obese women (Prentice et al., 1986)

Lean (n=13)	6.13	7.05	7.48	7.48	7.53	7.58
7.90	8.08	8.09	8.11	8.40	10.15	10.88
Obese(n=9)	8.79	9.19	9.21	9.68	9.69	9.97
11.51	11.85	12.79				

Null Hypothesis: The energy expenditure in Lean and Obese women is the same

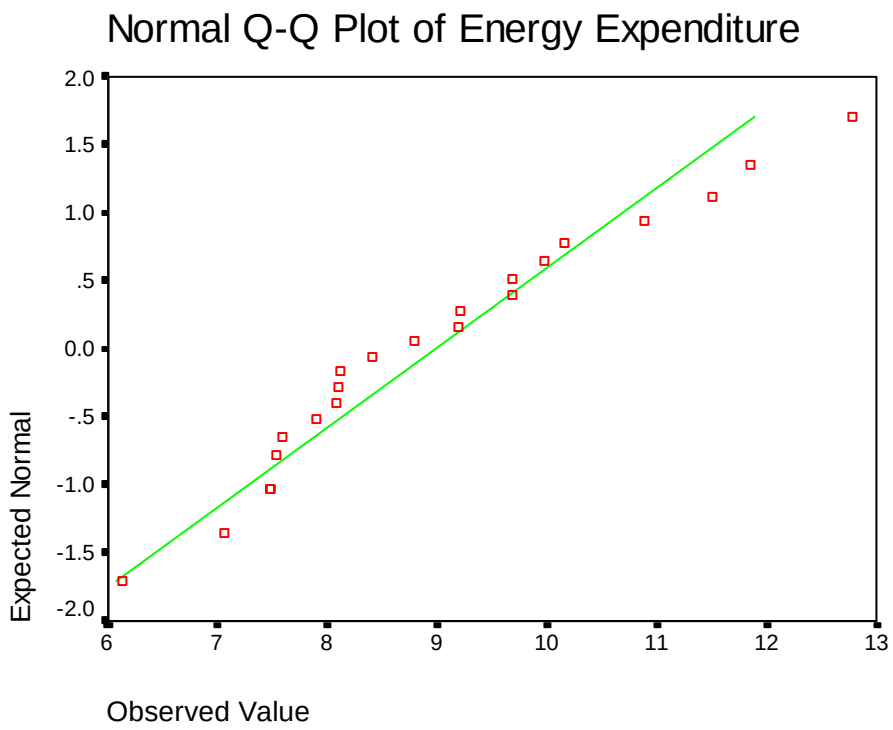
Alternate hypothesis: The energy expenditure in Lean and Obese women is different

First check for normality



Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Energy expenditure	.954	22	.381



Therefore we can use parametric testing

Independent Groups T-test

Group Statistics

Energy expenditure	N	Mean	Std. Deviation	Std. Error Mean
lean	13	8.0662	1.23808	.34338
obese	9	10.2978	1.39787	.46596

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Energy Expenditure	Equal variances assumed	1.002	.329	-3.946	20	.001	-2.2316	.56560	-3.41145	-1.05180
	Equal variances not assumed			-3.856	15.919	.001	-2.2316	.57882	-3.45917	-1.00408

The p-value for this test actually = 0.0008

This means that if the energy expenditure in Lean and Obese women is really the same, then the probability of obtaining the difference as shown by the sample (or a larger difference) is 0.0008 (or 8 in 10000).

This means that it is very unlikely and we can reject the null hypothesis of ‘no difference’

We can also note that the 95% confidence interval for the difference is (-3.41, -1.05) which does not include zero. This is an alternative way of confirming a significant difference at the 5% level.

Example 4

Snoring behavior in relation to presence or absence of heart disease (Norton and Dunn, 1985)

Heart Disease	Snoring				Total
	Non-snorers	Occasional snores	Nearly every night	Every night	
Yes	24	35	21	30	110
No	1355	603	192	224	2374
Total	1379	638	213	254	2484
% HD	1.7	5.5	9.9	11.8	

Null Hypothesis: Snoring is independent of HD

Alternate hypothesis: Snoring is not independent of HD (there is a trend - i.e. the more they snore the more likely they are to have HD)

We do a χ^2 (trend test) – also known as linear-by-linear association test

When we look at the numbers expected in each cell under the null hypothesis of no association/trend we can see that there are a lot less non-snorers and a lot more snorers with heart disease than would be expected if there was no association

Heart Disease	Snoring				Total
	Non-snorers	Occasional snores	Nearly every night	Every night	
Yes	24 (61.1)	35 (28.3)	21 (9.4)	30 (11.2)	110
No	1355 (1317.9)	603 (609.7)	192 (203.6)	224 (242.8)	2374
Total	1379	638	213	254	2484

P-value (p<0.0005)

(trend test) – also known as linear-by-linear association test

If there is no trend in the population, then probability of observing the above sample trend is negligible (almost impossible).

Therefore we can reject the null hypothesis of no trend and can conclude that it appears likely that those who snore more often are more likely to have HD (or vice versa!)

Example 5 – Times taken by children to complete patterns from the WISC scale

- The ‘row’ group were told to start with a row of three blocks and the corner group were asked to start with a corner of three blocks

Null Hypothesis: The average times taken in the two groups are the same

Alternate hypothesis: The average times taken in the two groups are different

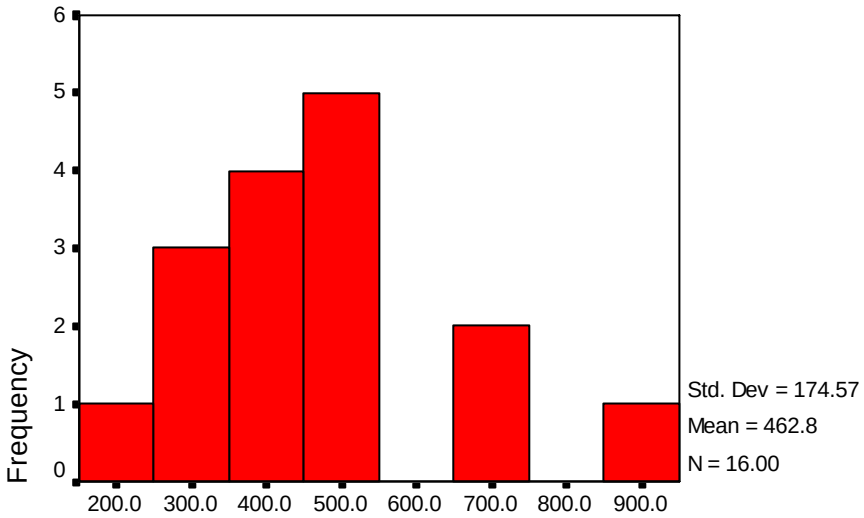
- First test for normality

			Time to completion
WISC Test Group	Row Group	1	675.00
		2	510.00
		3	490.00
		4	850.00
		5	317.00
		6	464.00
		7	525.00
		8	298.00
		9	491.00
		10	196.00
		11	268.00
		12	372.00
		13	370.00
		14	739.00
		15	430.00
		16	410.00
		Total N	16

			Time to completion
WISC Test Group	Corner Group	1	342.00
		2	222.00
		3	219.00
		4	513.00
		5	295.00
		6	285.00
		7	408.00
		8	543.00
		9	298.00
		10	493.00
		11	317.00
		12	407.00
		13	290.00
		14	301.00
		15	325.00
		16	360.00
		Total N	16

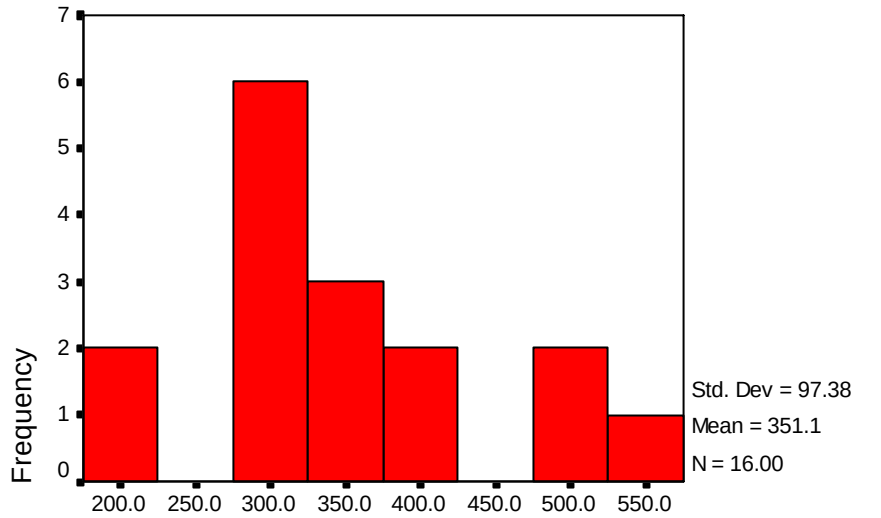
Histogram

For GROUP= Row Group



Histogram

For GROUP= Corner Group

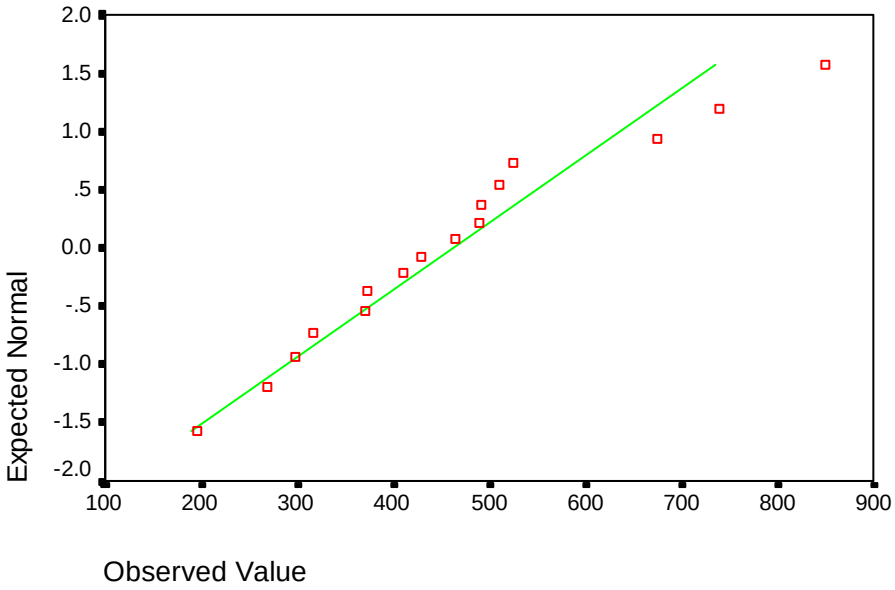


Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Time to completion	.910	32	.011

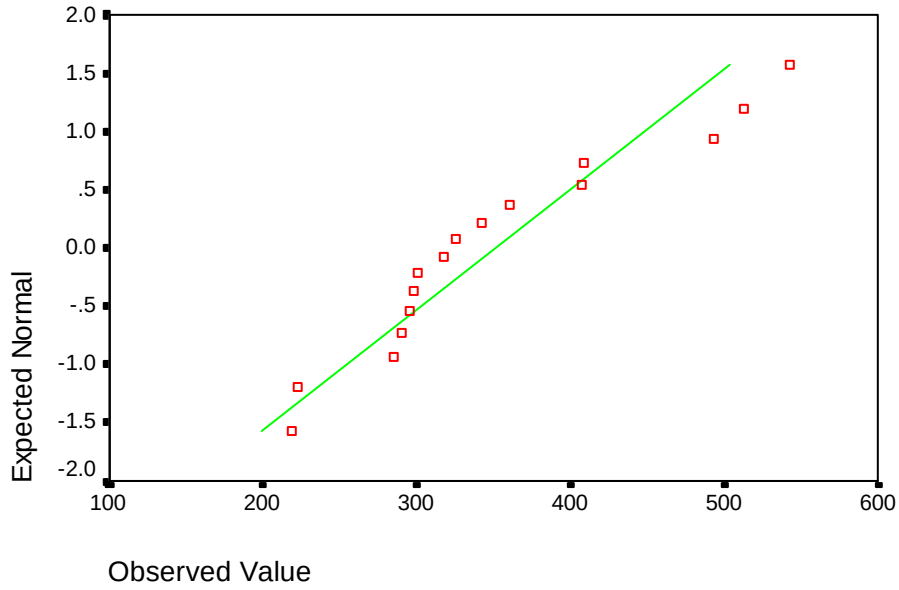
Normal Q-Q Plot of Time to completion

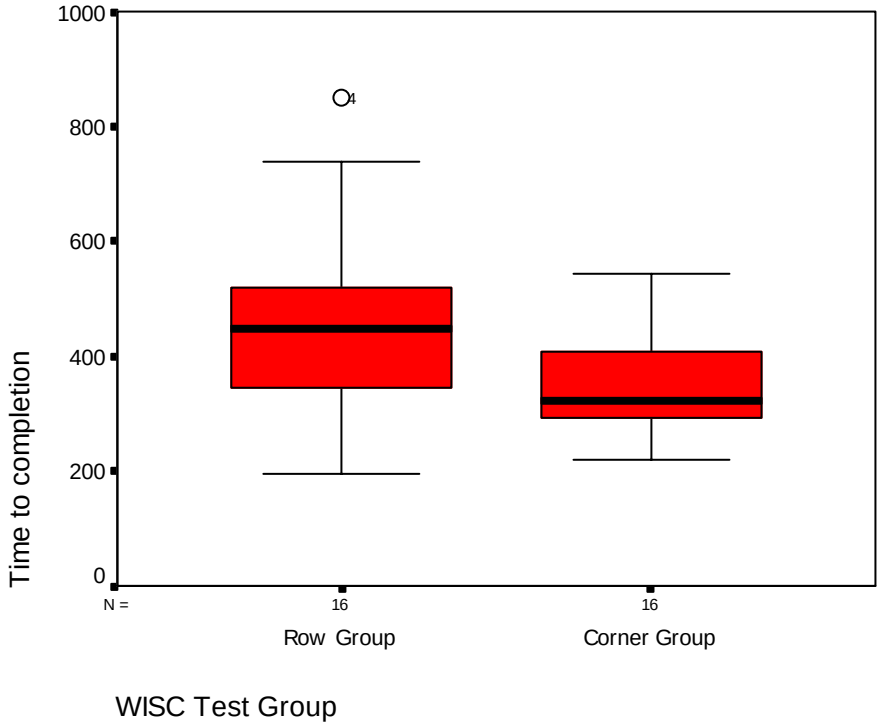
For GROUP= Row Group



Normal Q-Q Plot of Time to completion

For GROUP= Corner Group





- We decide that the distributions are not very normal according to the tests of normality and the shape of the histograms – both distributions appear very skew.
- We carry out a non-parametric test

  The Mann-Whitney U Test

Mann-Whitney Test

Ranks

	WISC Test Group	N	Mean Rank	Sum of Ranks
Time to completion	Row Group	16	19.69	315.00
	Corner Group	16	13.31	213.00
	Total	32		

Test Statistics[b]

	Time to completion
Mann-Whitney U	77.000
Wilcoxon W	213.000
Z	-1.922
Asymp. Sig. (2-tailed)	.055
Exact Sig. [2*(1-tailed Sig.)]	.056[a]

a. Not corrected for ties.

b. Grouping Variable: WISC Test Group

Percentiles

			Percentiles		
		WISC Test Group	25	50	75
Weighted Average(Definition 1)	Time to completion	Row Group	330.2500	447.0000	521.2500
		Corner Group	291.2500	321.0000	407.7500

Conclusions

- We cannot reject the null hypothesis of no difference between the two groups but the result is very close to significance. It would be worth checking the data to make sure it was recorded correctly. At least one of the groups appears to have an outlier.
- The result is still worth reporting even if this outlier is found to be correct.
- It might be that the study did not have enough power to find a significant difference (at p=0.05) so a larger study could be planned with sufficient numbers to ensure adequate power. The sample size for this new study could be calculated using statistics obtained from this smaller study.

Reminder

Sample Size Calculations

- Why perform sample size calculations?
  - We need to know how many people are required to detect a treatment difference should it exist. If a study is too small, it may lack the *power* to pick up differences.
  - Clinical trials can be expensive – we don't want to waste time, money and resources on a study that will unable to pick up clinically relevant differences. We also don't want to study more people than necessary.
  - It is unethical to subject patients to treatment unnecessarily – either by performing a trial with low power or by recruiting more patients than necessary.

Statistical Consulting Unit (SCU)

- Provides Statistical Consultancy services internally (within UL) and externally (e.g. to HSE Mid-Wesrern Area Staff)

- Provides Basic Courses in quantitative research methods internally and externally on a regular basis

Overview of SCU - Courses currently offered

- **Courses are offered several times a year**
- **Jan/Feb and June/July each year**
- **Next set of courses**

Early June 2005

June 7th Questionnaire Design

June 8th Surveys and Sampling

June 14th SPSS for Beginners

Exploratory Data Analysis

Further Data Analysis

June 15th – 16th Basic Statistics for Researchers

Introductory Design of Experiments

Overview of SCU - Future Courses

- Introductory course for clinical researchers on pro-forma design, database construction in excel and importing into SPSS, Data Manipulation plus simple summary statistics and graphics in SPSS
- Introductory PK/PD course Aug 24th -25th 2005
- Logistic Regression
- Other 'basic' courses that might be useful to researchers
- Any other 'further' courses requested that have sufficient demand

Overview of SCU – Summary

1. **Contact the SCU as early as possible in a study**
2. **Provide as much information as you can**
3. **If contact not made early (for whatever reason!) the SCU is still happy to get involved at any stage of study and give any advice needed**
4. **Courses are available to consolidate knowledge of quantitative methods and use of statistical software**
5. **Statisticians are friendly people - honest!**

SCU – Contact Details

Dr Jean Saunders
Executive Director
Statistical Consulting Unit
(Office of the Vice President Academic & Registrar)
Department of Mathematics and Statistics
University of Limerick
Tel: +353 - 61 – 213471
Mob: +353 – 86 - 3866353
Fax: +353 - 61 - 334927
email jean.saunders@ul.ie