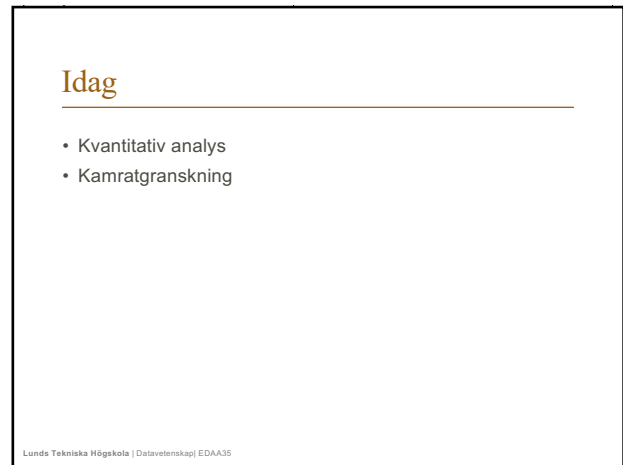
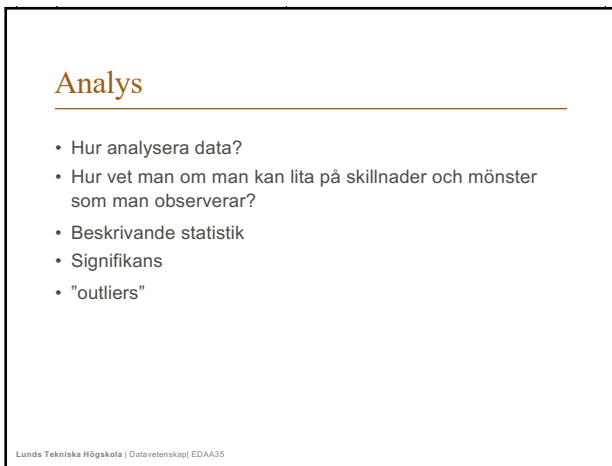


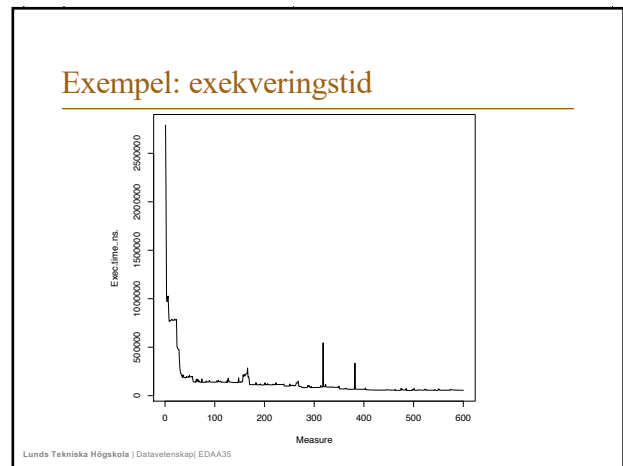
1



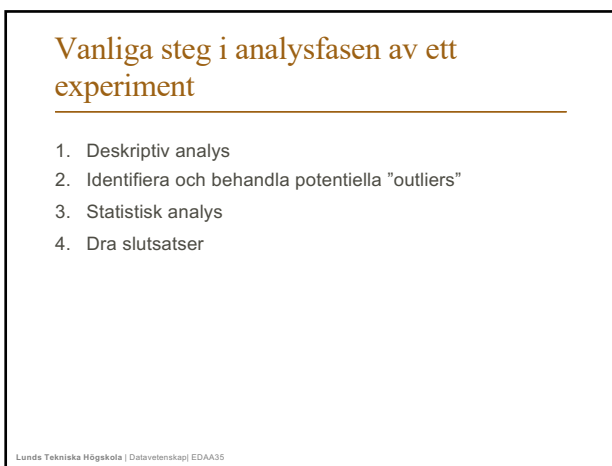
2



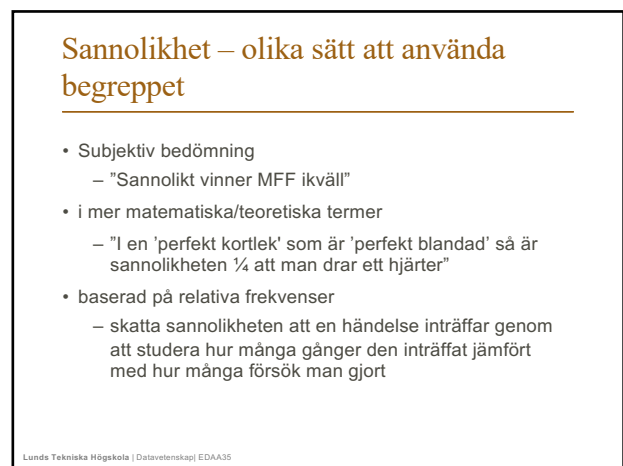
3



4



5



6

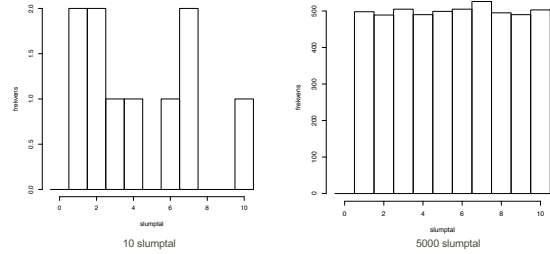
forts

- Beskrivs som ett värde mellan 0 och 1
 - 0: det händer aldrig
 - 1: det händer varje gång
- Ex: $P(\text{klave}) = 1/2$

Lunds Tekniska Högskola | Datavetenskap | EDAA35

7

Histogram för slumptal

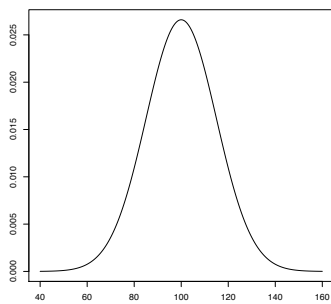


...det verkar finnas en teoretisk fördelning bakom de observationer vi gör...

Lunds Tekniska Högskola | Datavetenskap | EDAA35

8

Normalfördelning, täthetsfunktion



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Lunds Tekniska Högskola | Datavetenskap | EDAA35

9

Deskriptiv analys

Ex

- Centralitet
- Spridning
- ...men också grafisk analys som diagram etc

Lunds Tekniska Högskola | Datavetenskap | EDAA35

10

Centralitet

- Medelvärde** $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- Median**, mittersta värdet (eller medelvärdet av de två mittersta om det är ett jämnt antal värden)
- Typvärde** / modalvärde / "mode": vanligaste värdet
- Exempel-data: 1, 2, 3, 4, 20
- Medelvärde: 6
- Median 3

Lunds Tekniska Högskola | Datavetenskap | EDAA35

11

Spridningsmått

- Standardavvikelse $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
- Variationsvidd $R = \max_i x_i - \min_i x_i$

Lunds Tekniska Högskola | Datavetenskap | EDAA35

12

Skalor (repetition)

- Nominell
- Ordinal
- Intervall
- Ratio

Lunds Tekniska Högskola | Datavetenskap | EDAA35

13

När bör man använda vilket mått?

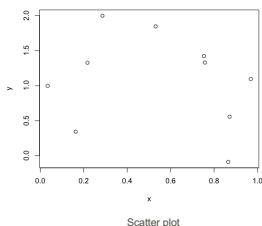
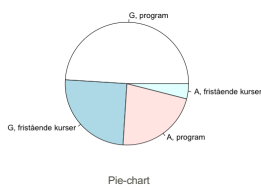
Mått / Skala	Nominell	Ordinal	Intervall	Ratio
Medelvärde			ok	ok
Median		ok	ok	ok
Typvärde	ok			
Standardavvikelse			ok	ok
Variationsvidd		ok	ok	ok

Lunds Tekniska Högskola | Datavetenskap | EDAA35

14

Grafisk presentation, två exempel

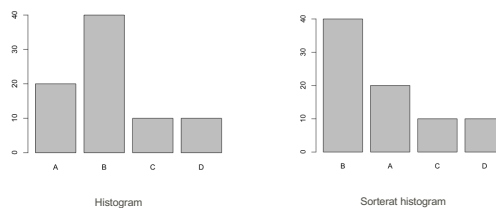
LU, studenternas fördelning på utbildningsnivå



Lunds Tekniska Högskola | Datavetenskap | EDAA35

15

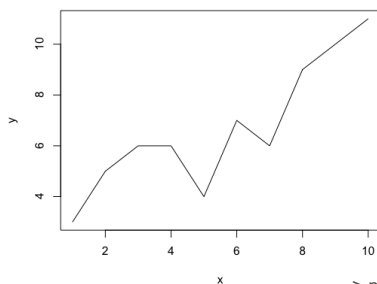
Fler alternativ



Lunds Tekniska Högskola | Datavetenskap | EDAA35

16

”Vanlig” plot



Lunds Tekniska Högskola | Datavetenskap | EDAA35

17

```

plot (graphics)
Generic X-Y Plotting
R Documentation

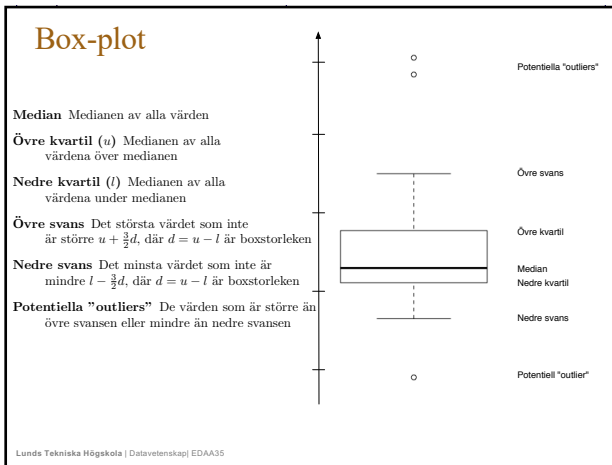
Description
Generic function for plotting of R objects. For more details about the graphical parameter arguments, see par.
For simple scatter plots, plot.default() will be used. However, there are plot methods for many R objects, including
factor, data.frame, matrix objects, etc. Use methods(plot) and the documentation for these.

Usage
plot(x, y, ...)

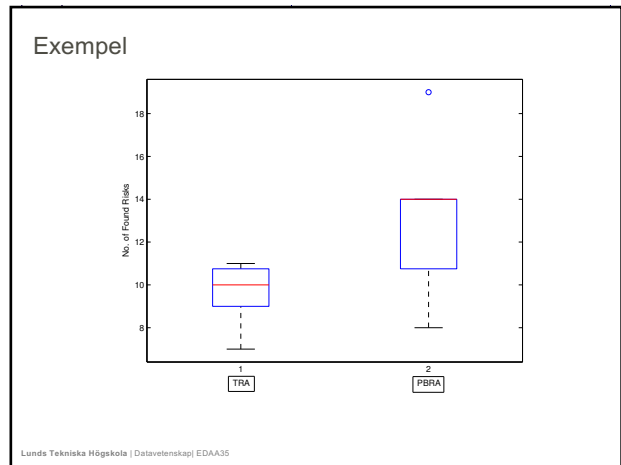
Arguments
x
the coordinates of points in the plot. Alternatively, a single plotting structure, function or any R object with a
plot method can be provided.
y
the y coordinates of points in the plot, optional if x is an appropriate structure.
...
Arguments to be passed to methods, such as graphical parameters (see par). Many methods will accept the
following arguments:
type
what type of plot should be drawn. Possible types are
• "p" for points,
• "l" for lines,
• "b" for both,
• "c" for the lines part alone of "b",
• "o" for both "overplotted",
• "h" for "histogram" like (or "high-density") vertical lines,
• "s" for stair steps.
    
```

Lunds Tekniska Högskola | Datavetenskap | EDAA35

18



19



20

"Outliers"

Identifiera

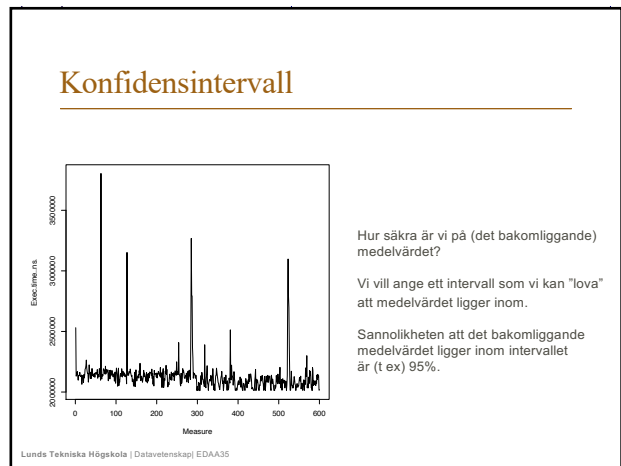
- Boxplot
- Scatter plot
- Kvalitet på datapunkt, tex anda frågor?
- ...

Vad göra?

- Ta bort?
- Behålla?
- Analysera separat?

Lunds Tekniska Högskola | Datavetenskap| EDAA35

21



22

Exempel med påhittade observationer

```

confidenceInterval <- function(x, confidenceLevel=0.95)

> source("/usr/local/cs/EDAA35/R_resources.R")
> x <- c(4, 5, 4, 4)
> confidenceInterval(x)
lower upper
3.454388 5.045612
> x <- c(4, 5, 4, 4, 4, 5, 4, 4, 4, 5, 4, 4, 4, 4, 5, 4, 4)
> confidenceInterval(x)
lower upper
4.011697 4.488303
>
    
```

Lunds Tekniska Högskola | Datavetenskap| EDAA35

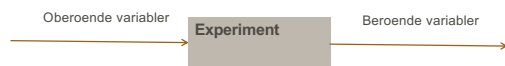
23

- ### forts
- Ju fler datapunkter desto mindre intervall
 - Ju större underliggande spridning desto större intervall
 - Ju större krav på säkerhet (t ex 95%) desto större intervall
 - Bygger på ett antal antaganden! (se statistikkursen)
- Lunds Tekniska Högskola | Datavetenskap| EDAA35

24

Jämförande studie - Experiment

- Jämförande studie där man vill ta reda på effekten av ett antal variabler på ett antal andra variabler



Lunds Tekniska Högskola | Datavetenskap | EDAA35

25

Exempel

- Undersöka effekten av checklista vid granskning

Person	Teknik	Antal fel	Tid (min)
A	Checklista	34	87
B	Ad-hoc	22	121
C	Checklista	43	90
D	Ad-hoc	30	87
E	Checklista	35	81
F	Ad-hoc	21	99
G	Checklista	45	104
H	Ad-hoc	39	78
I	Checklista	32	89
J	Ad-hoc	32	66

Lunds Tekn

26

Exempel 2

- Jämföra hur lång tid två olika implementationer tar att exekvera

Körning	Implementation	Tid (ms)
1	A	145
2	B	122
3	A	155
4	B	145
5	A	150
6	B	156
7	A	145
8	B	151
9	A	144
10	B	143

Lunds Tekniska Högi

27

Experiment – viktiga aktiviteter

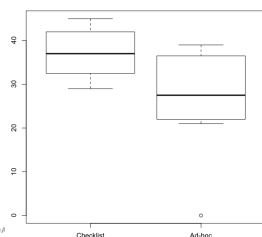
- Definition av population, sampling (när det är relevant)
- Randomisering – dela upp i grupper
- Genomför, gör mätningar
- Analysera mätningar
- Dra slutsatser

Lunds Tekniska Högskola | Datavetenskap | EDAA35

28

Exempel (ad-hoc – checklista) forts

- Data (hypotetisk exempeldata)
 - Checklist: 34, 43, 35, 45, 32, 44, 29, 39, 41, 29, 33, 39
 - Ad-hoc: 22, 30, 21, 39, 32, 36, 37, 38, 22, 23, 25, 0

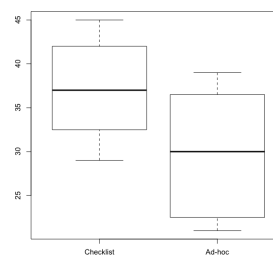


Lunds Tekniska Högskola | Datavetenskap | ED

29

forts

- En granskning genomfördes felaktigt → ta bort den



Vad kan vi säga nu?
Kan vi 'lova' att checklista är bra?

Lunds Tekniska Högskola | Datavetenskap | EDAA35

30

T-test

```
> t.test(Checklist, Adhoc)
```

Welch Two Sample t-test

data: Checklist and Adhoc

t = 2.7156, df = 19.07, p-value = 0.01369

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

1.691264 13.051160

sample estimates:

mean of x mean of y

36.91667 29.54545

Lunds Tekniska Högskola | Datavetenskap | EDAA35

31

T-test, några viktiga begrepp

- H_0 : "nollhypotes" – "ingen skillnad" (underliggande medelvärdena är samma)
- H_a : Alternativ hypotes
- Testet går ut på att se om man kan "förkasta" H_0
- Typ-I-fel: P(förkasta H_0 även om H_0 är sann)
- Typ-II-fel: P(inte förkasta H_0 även om H_0 är falsk)
- Bygger på att antal antagande! (se statistikkurs)
- Visa i terminalen

Lunds Tekniska Högskola | Datavetenskap | EDAA35

32

Några exempel på andra test

Design	Parametriska	Icke-parametriska
En oberoende variabel med två nivåer, varje person gör en sak	t-test	Mann-Whitney
En oberoende variabel med två nivåer, varje person gör båda sakerna	"paired t-test"	Wilcoxon
En oberoende variabel med mer än två nivåer	ANOVA	Kruskal-Wallis
Mer än en oberoende variabel	ANOVA	

...viktigt att förstå under vilka förutsättningar olika test kan användas!

Lunds Tekniska Högskola | Datavetenskap | EDAA35

33

Kamratgranskning



34

Kamratgranskning

- På lab 5
- Varje rapport granskas av tre andra studenter
- Alla ska uppdatera enligt granskningskommentarerna
 - Enligt alla kommentarer som stämmer (och är möjliga att ändra enligt)
 - Bra att sammanfatta hur man tagit hand om kommentarerna

Lunds Tekniska Högskola | Datavetenskap | EDAA35

35

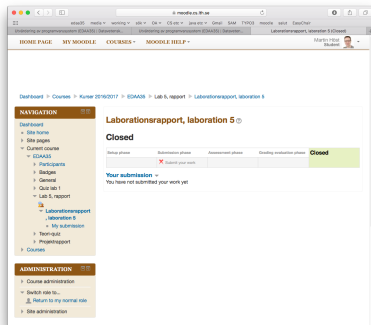
Granskningsmall

- Är målsättningen med den rapporterade studien tydlig?
- Är den övergripande metoden som används tydligt beskriven och motiverad?
- Är det tydligt vilken data som samlats in och hur den har samlats in?
- Presenteras tillräckligt mycket av den data som samlats in?
- Är analysmetoderna rätt valda och lämpliga för studien?
- Är analysen korrekt genomförd och kan man lita på resultaten?
- Finns det en beskrivning av hur validitetshot tagits om hand och en diskussion om eventuella kvarvarande validitetshot?
- Är slutsatserna tydliga och svarar de på den målsättning som finns med studien?
- Är språket lätt att förstå och av tillräckligt hög kvalitet?
- Följer rapporten en lämplig struktur?
- Finns det numererade figurtexter till alla figurer och numererade tabelltexter till alla tabeller, samt är alla figurer och tabeller refererade i texten?

Lunds Tekniska Högskola | Datavetenskap | EDAA35

36

Hur? (I Moodle)



Lunds Tekniska Högskola | Datavetenskap | EDAA35

37

Varför?

1. Man lär sig mycket genom att granska
2. Man får många kommentarer

Lunds Tekniska Högskola | Datavetenskap | EDAA35

38