**Brev till Block 3 och 4**

Hej!

Här kommer ett brev som berättar om tentan och kommenterar teori för block 3. Du kommer att hitta läshänvisningar för ISwR och Lind i brevet

**Så här hoppas vi att kursen fungerar:**

**R**

Du lär dig R genom att räkna praktiska exempel i stigande svårighetsgrad.

**Statistik**

Du ökar din kunskap om statistik genom att läsa i Lind, vägledd av **ditt** behov.

Du använder R för att utforska statistiska koncept och metoder i gruppuppgifterna.

Du använder gruppuppgifter och ISwR som kokbok för att lösa liknande problem i hemtentan.

Motto: Svåra gruppuppgifter – lätt individuell tenta.

**Idén med individuell uppgift och individuell tentamen**

För att få betyg måste du också visa upp vad du kan göra på egen hand. Här vill vi kontrollera din förmåga att lösa statistiska problem med R med gruppuppgifter, eller ISwR, eller kanske internet som recept. Du kommer att få genomföra momentet ”Självtest”, ett test i Canvas. Deadline framgår av schemat. Tanken med testet är att alla får se hur den riktiga tentan fungerar rent tekniskt.

**Hur gör man för att få VG på kursen?**

Samtidigt som tentan öppnar publiceras en VG-tentauppgift. Svaren ska du maila till Eric (eric.rullman@ki.se) inom tentans dygn i form av **körbar, kommenterad R-kod**. Dessutom behöver du få godkänt på block5, deadline se schemat, en VG-uppgift som du kan lösa i grupp eller på egen hand.

**Hur ska man förbereda sig för tentan?**

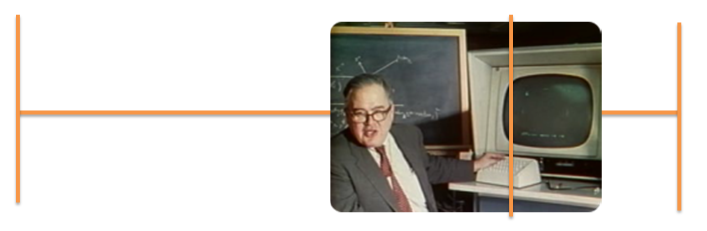
Gå igenom de tre gruppuppgifterna 2, 3 och 4. Vi hoppas att du ska kunna använda dig av gruppuppgifterna, ISwR och en statistikbok för att ta dig an problem i statistisk programmering du stöter på efter kursen. I tentan vill vi testa den förmågan. Man ska kunna få G på tentan genom att hitta bra i sina gruppuppgifter och i boken.

**Kommentarer om block 3**

Uppgift 3 handlar först om deskriptiv statistik. Hur kan man sammanfatta en stor datamängd? Centralmått och variationsmått är två siffror som kan fungera, rätt använda. En rad olika grafer kan också användas. Box-plot är en graf som sammanfattar en datamängd med 5 siffror: Lägsta värdet och de fyra kvartilerna.

Vem hittade på box-plot?

Thank you Dr Tukey for the box-plot. Exploratory Data Analysis, 1977.



Efter deskriptiv statistik ger vi oss på statistiska fördelningar med bionomialfördelning och normalfördelning som exempel. Ni får ett par praktiska räkneexempel som ska visa hur man kan använda fördelningar.

Lägg märke till ett grundläggande koncept:

Arean under kurvan, vilken fördelning det månde vara, är 1. Det innebär att hela populationen ryms här, eller alla möjliga utfall. Genom att mäta delareor kan du beräkna hur ovanligt ett utfall är.

**Genomräknat exempel**

Fasteblodsockernivån är normalfördelad med medelvärdet 4,6 mmol/l och standardavvikelsen 1,1 mmol/l. Enligt WHO definieras diagnosen diabetes som fasteblodsockervärde större än 6,7 mmol/l. Hur stor del av befolkningen skulle klassificeras som diabetiker enligt denna definition?

Vi har ritat en normalfördelning nedan. Arean under hela kurvan är 1, dvs den representerar hela populationen. Sedan har vi ritat ett vertikalt streck vid 6.7 mmol/l. Arean till höger om strecket, som vi färgat blå, representerar andelen av befolkningen som har ett fasteblodsockervärde över 6.7.

Klistra in denna kodrad i R console:

1 - pnorm(6.7, 4.6, 1.1) # andelen personer som har 6.7 eller högre

2.8% av befolkningen klassas som diabetiker enligt definitionen med fasteblodsockervärde.



**Varför är ”allting” normalfördelat?**

Det kan verka ointuitivt att många orelaterade variabler inom biologi, kemi, fysik skulle följa samma mystiska fördelning med denna ogenomträngliga formel:

C:\Users\Andreas\Documents\KI\undervisning\StatistiskaMetoderMedR-2012VT\SMMR2-veckobrev\gauss.png

I block 3 finns två övningar som inte har karaktären av bevis, men som illustrerar användbarheten av normalfördelningen. Den ena visar att normalfördelningen under vissa omständigheter kan vara en god approximation av binomialfördelningen. Den är betydligt lättare att förstå sig på och relatera till naturliga fenomen. Den andra övningen illustrerar centrala gränssnittssatsen. Läs för all del Lind 8.5 Central Limit Theorem, eller en annan bok eller Wikipedia.

**Läshänvisningar ISwR för block 3**

Använd kapitel tre och fyra

**Lärandemål ur Lind 15ed – läs om du vill**

Jag tycker om upplägget i senaste upplagorna av Lind med lärandemål som markeras i texten. Jag har valt ut ett antal lärandemål i Lind som är relaterade till block 3. Se läromålen nedan som ett sätt att hitta mellan inlämningsuppgiften och Lind*.* ***Du kan lära dig hantera R med hjälp av övningsuppgifterna utan att läsa Lind och det är helt OK för den här kursen.*** Du kan också ta tillfället i akt och läsa mycket statistik för att få ut max av uppgifterna. Börja med att läsa igenom lärandemålen och kolla vilka du känner igen sedan tidigare.

Andra upplagor av Lind skiljer sig lite i numrering och formulering av vissa lärandemål, men det spelar ingen roll vilken upplaga du använder.

**Listan nedan återknyter till koncept du möter i gruppuppgifterna.**

Ch1, p1 Lind 15ed läs om du vill

2 Know the differences between descriptive and inferential statistics.

3 Understand the difference between a sample and a population.

4 Explain the difference between qualitative and quantitative variables.

5 Compare the differences between discrete and continuous variables.

6 Recognize the levels if measurement in data.

Ch2, p21 Lind 15 ed läs om du vill

1 Make a frequency table for a set of data.

2 Organize data into a bar chart.

3 Present a set of data in a pie chart.

4 Create a frequency distribution for a data set.

5 Understand a relative frequency distribution.

6 Present data from a frequency distribution in a histogram or a frequency polygon.

Ch3, p57 Lind 15 ed läs om du vill

1 Explain the concept of central tendency.

2 Identify and compute the arithmetic mean.

4 Determine the median.

5 Identify the mode.

7 Explain and apply measures of dispersion.

Ch4, p102 Lind 15 ed läs om du vill

1 Construct and interpret a dot plot.

3 Identify and compute measures of position.

4 Construct and analyze a box plot.

6 Create and interpret a scatter diagram.

Ch5, p144 Lind 15 ed läs om du vill

1 Explain the terms experiment, event and outcome.

Ch6, p186 Lind 15 ed läs om du vill

1 Identify the characteristics of a probability distribution.

2 Distinguish between a discrete and a continuous random variable.

3 Compute the mean of a probability distribution.

5 Describe and compute probabilities for a binomial distribution.

Ch7, p222 Lind 15 ed läs om du vill

1 List the characteristics of the uniform distribution.

3 List the characteristics of the normal distribution.

4 Convert a normal distribution to the standard normal distribution.

5 Find the probability that a normally distributed variable is between two values.

7 Approximate the binomial distribution using the normal distribution.

Ch8, p265 Lind 15 ed läs om du vill

4 Describe the sampling distribution of the sample mean.

5 Explain the central limit theorem.

7 Apply the central limit theorem to find probabilities of selecting possible sample means from a specified population.

Hälsningar

Mirko och Eric