

Övningar till GeoGebra

Datorövning L9MA30, LGMA30, Matematisk statistik, Hösten 16

26 september 2017

Examinator: Claes Andersson, mejl andclae@chalmers.se, tel. 031 772 8294

Datorövningen är obligatorisk. Om man inte kan delta datorövningen, måste en rapport (3-8 sidor) med lösningsförslag, resultat och diagram (med tolkning) lämnas in senast på måndag, den 2 oktober 2017.

Extra uppgiftet (ej obligatoriskt) ger maximalt 0.5 bonus poäng till tentamen (krävs ≥ 2 poäng). Sista inlämningsdag för extra uppgiftet är måndag, den 2 oktober 2017.

I denna datorbaserade övning skall ni bekanta er med programvaran GeoGebra, som är gratis att ladda ner från www.geogebra.org. Börja med att starta programmet, (byt språk om det behövs och starta om) välj perspektiv 'Sannolikhet' och lägg till 'Kalkylblad' som du hittar under 'Visa'.

1) Den klassiska sannolikhetsdefinitionen

Simulera diskreta slumpetal av $U(10)$ -fördelningen. Detta gör ni genom att i kalkylbladet skriva

$$=SlumptalMellan[1, 10].$$

Markera cellen och dra neråt för flera slumpetal. Prova att variera antalet observationer, t.ex. $n = 10, 50, 100$. M.h.a. knappen 'envariabelanalys' kan du få fram medelvärde, histogram o.s.w.. Vilken är den empiriska sannolikheten för värdet $\mathbb{P}(3 \leq X < 7)$? Jämför med den teoretiska sannolikheten.

2) Jämförelse av fördelningar

Du kan titta på väntevärde, varians, fördelnings- eller sannolikhetsfunktion o.s.v..

• **Binomial- och Poissonfördelning**

Jämför en $Po(3)$ med en $Bin(10, 0.3)$, $Bin(100, 0.03)$ fördelning. Vad ser du för mönster? Använd t.ex.

$$=SlumptalBinomialfördelat[10, 0.3].$$

och

$$=SlumptalPoissonfördelat[3].$$

• **Binomial- och Normalfördelning**

1. Jämför en $Bin(n, p)$ med en $N(\mu = np, \sigma = \sqrt{np(1-p)})$ fördelning. Använd

$$=SlumptalNormalfördelat[\mu, \sigma].$$

2. För vilka val av n och p är Normalfördelningen en bra approximation? När är den olämplig? Ledning: Hur ser kvantildiagrammen ut för olika Binomialfördelningarna med varierande värden för n och p ?

3) Statistisk analys

Data för denna uppgift hittar du på hemsidan. Excel filen heter *dataGeogebra* och har tre ark, ett för varje uppgift nedan. För att importera data till Geogebra kan du enkelt kopiera de relevanta cellerna och klippa in dem i Geogebbras kalkylblad.

- **Envariabel analys**

Anna samlar in data från antalet bilar som passerar vid de vägar där vägtullar är placerat ut i Göteborg.

1. Gör en skattning av väntevärde och varians.
2. Illustrera stickprovet m.h.a. histogram. Hur kunde antalet bilar vara fördelat (om man antar kontinuerliga utfall)?

- **Tvåvariabel regressionsanalys**

Staffan har blivit tillfrågad att hjälpa till med en studie för att undersöka sambandet mellan viktnedgång genom ett kirurgiskt ingrepp och dess effekt på blodtrycket. I en pilot-studie har de ansvariga för studien fått fram resultat från 30 deltagare.

1. Illustrera observationerna i en två-dimensionell figur med skillnad i BMI på x-axeln och skillnad i blodtryck på y-axeln. Verkar det finnas ett linjärt samband mellan x och y (ligger värdena samlade runt en tänkt linje i spridningsdiagrammet)?
2. Vad är korrelationen mellan x och y? Beskriv sambandet i termer av frågeställningen.

- **Flervariabel analys**

Erik vill ta reda på hur vilopulsen påverkas av meditation. För att göra detta frågar han 15 av sina vänner om de vill delta i hans 'studie'. Han mäter deras vilopuls i början av studien och efter 10 dagar med 30 min meditation/dag.

1. Illustrera materialet grafiskt och beräkna variationsbredderna.
2. Jämför de empiriska fördelningar före och efter 10 dagar med meditation. Tror du att meditationen hjälper att sänker vilopulsen?

4) Extra uppgift (ej obligatorisk och GeoGebra behövs inte, men får användas)

En resebyrå undrar om antal sålda resor x och omsättning y (i 10000 kr) är positivt linjärt korrelerade under de senaste sju åren:

i	År	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2009	200	470	94000	40000	220900
2	2010	250	630	157500	62500	396900
3	2011	300	640	192000	90000	409600
4	2012	280	630	176400	78400	396900
5	2013	350	790	276500	122500	624100
6	2014	400	930	372000	160000	864900
7	2015	425	950	403750	180625	902500
Summan		2205	5040	1672150	734025	3815800

- a) Beräkna och tolka korrelationskoefficienten $\rho = \frac{s_{xy}^2}{s_x \cdot s_y}$. Är det rimligt att anta ett (statistiskt signifikant) linjärt samband mellan x och y ?

2p

- b) Antag att det finns ett linjärt samband mellan x och y , så att $y_i = a + bx_i$. Vilken omsättning kan resebyrån förvänta sig i 2016 om de säljer 500 resor detta år?

2p