

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 10 juni 2019, kl. 8.00 – 12.00.

Kursens förväntade läranderesultat enligt kursplanen

Efter genomförd kurs skall du kunna:

1. analysera fördelningen hos en datamängd avseende centralvärde och spridning, såsom medelvärde och median samt standardavvikelse, samt visualisera detta.
 2. redogöra för olika synsätt på begreppet sannolikhet.
 3. beräkna sannolikheter för händelser, med användning av begrepp och verktyg som oberoende, betingning, oförenlighet, komplementhändelse, union, snitt, kombinatorik, trädidiagram.
 4. formulera en sannolikhetsmodell med hjälp av stokastiska variabler, även med centrala gränsvärdesatsen, och använda den för att bestämma egenskaper hos dess fördelning samt beräkna sannolikheter.
 5. beräkna punktskattningar av väntevärde, varians, standardavvikelse, sannolikhet och intensitet samt bedöma deras lämplighet.
 6. beräkna konfidensintervall för väntevärde (med och utan känd standardavvikelse), sannolikhet och intensitet samt tolka resultatet.
 7. formulera och genomföra en hypotesprövning, och däri kunna tolka begreppen styrkefunktion och p -värde.
 8. genomföra en korrelationsanalys och tolka resultatet.
 9. ställa upp och tolka en linjär regressionsmodell med två variabler, avgöra om en linjär modell är tillämpbar, samt bedöma tillförlitligheten hos skattningar av såväl väntevärden som enskilda observationer.
 10. använda datorstöd för alla beräkningar där det är relevant.
-

Tillåtna hjälpmedel:

- Valfri bok inom statistik och sannolikhetslära¹
- Miniräknare av valfritt slag

Det får finnas anteckningar och markeringar i boken, inklusive "pagemarkers" (några centimeter stora), men inga lösblad.

Frågor besvaras av Michael Hörnquist som besöker salen minst två gånger under skrivtiden, ca kl. 9 och 10.30. Svar och kortfattade lösningsförslag finns på Studieinfo efter skrivningens slut. Skrivningsresultat finns i Ladok senast 15 arbetsdagar efter tentamenstillfället.

Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs automatiskt och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng.

Redovisningen av uppgifterna skall vara sådan att det klart framgår vad du gör. Det innebär att förklarande text skall finnas med. Att enbart ge ett svar utan motivering, även om det är korrekt, är normalt inte tillräckligt för att uppgiften skall anses vara nöjaktigt löst. Å andra sidan behöver inte ett felaktigt svar, såvida det inte av någon orsak är orimligt, innebära att lösningen ej är godkänd. Väsentligt är att det bakomliggande resonemanget klart framgår samt är relevant.

Lycka till!

¹Kurslitteraturen "Tillämpad statistik – en grundkurs", Wahlin, Sanoma förlag, torde vara vanligast.

1. Vi har i huvudsak kommit i kontakt med två mått för "mittenvärde", medelvärde och median, och två mått för spridning, populationsstandardavvikelse och stickprovsstandardavvikelse. Ge exempel på:
 - (a) en datamängd där median är större än medelvärde, eller visa att det inte finns en sådan mängd.
 - (b) en datamängd där medelvärde är större än median, eller visa att det inte finns en sådan mängd.
 - (c) en datamängd där populationsstandardavvikelsen är större än stickprovsstandardavvikelsen, eller visa att det inte finns en sådan mängd.
 - (d) en datamängd där stickprovsstandardavvikelsen är större än populationsstandardavvikelsen, eller visa att det inte finns en sådan mängd.

Respektive datamängd ska innehålla minst tre element.

2. Göran och Lisa är på hemväg från restaurangen Den gyllene moroten. På vägen passerar de nio trafikljus, och eftersom de är mätta och belåtna går de så långsamt att trafikljusen kan uppfattas som helt oberoende av varandra. Sannolikheten att ett enskilt trafikljus visar "grön gubbe" är 40%.
 - (a) Hur sannolikt är det att alla nio trafikljusen visar grönt?
 - (b) Vad är sannolikheten att de är så lyckosamma att sju eller fler av trafikljusen visar grönt?
 - (c) Om nu de första sju trafikljusen har visat grönt, vad är då sannolikheten att även de två sista gör det?
3. Vid ett järnverk produceras armeringsjärn som är avsedda att väga 2,0 kg, det så kallade riktvärdet. Vikten för ett enskilt armeringsjärn kan beskrivas med en normalfördelad slumpvariabel med $\mu = 2,0$ kg och $\sigma = 0,2$ kg.
 - (a) De armeringsjärn som avviker med mer än 0,25 kg från riktvärdet måste kasseras. Hur stor andel är det?
 - (b) Vad är sannolikheten att ett slumpmässigt valt armeringsjärn väger "helt rätt", tolkat som att dess faktiska vikt inte avviker mer än 0,1% från riktvärdet?

4. Opinionsinstitutet Novus februariundersökning gjordes under perioden 14 januari – 3 februari 2019. Man tillfrågade totalt 6006 personer, varav 3045 svarade. Deras bortfallsanalyser pekar på att bortfallet är åsiktsmässigt slumpmässigt utifrån undersökt målgrupp. Resultatet blev enligt tabellen nedan:

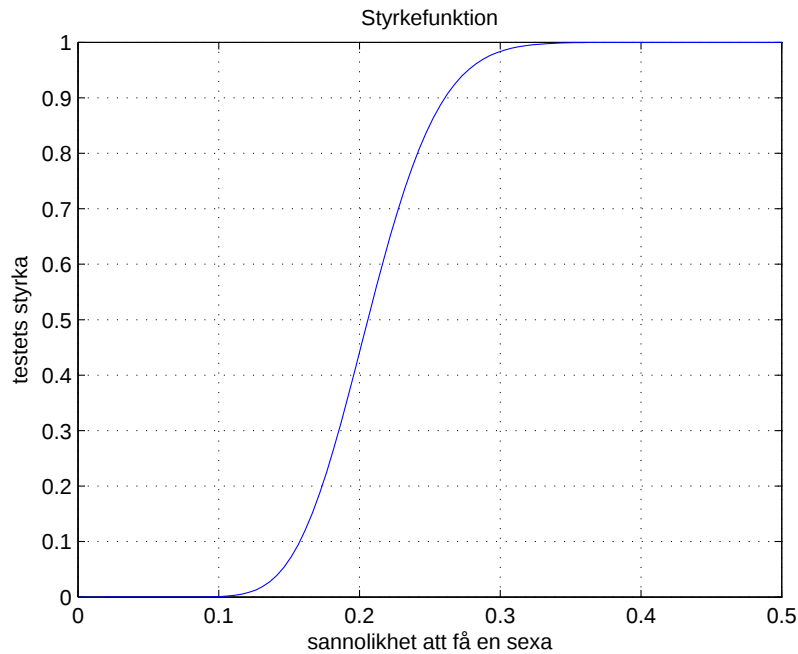
	Valet 2010	Valet 2014	22-aug- 18	30-aug- 18	07-sep- 18	Valet 2018	18- okt	18- nov	18- dec	19- jan	19- feb
M	30,1	23,3	19,3	19,4	17,7	19,8	19,7	19,2	18,3	18,9	18,3
L	7,1	5,4	5,3	6,2	6,5	5,5	4,6	3,7	3,9	3,4	2,7
C	6,6	6,1	10	8	8,6	8,6	8,4	8,9	8,6	7,6	7,5
KD	5,6	4,6	3,5	5,3	5,7	6,3	6	5,9	6,5	6,2	7,4
S	30,7	31	25,1	24,3	24,9	28,3	28,3	29,8	30,0	31,8	29,0
V	5,6	5,7	8,7	10,1	10	8	8,4	7,9	7,7	7,6	9,1
MP	7,3	6,9	6,2	5,7	5	4,4	4,4	4,3	4,0	3,9	3,4
SD	5,7	12,9	19,2	18,2	19,1	17,5	18,8	19,2	19,8	19,2	21,2
Annat parti	1,8	4,1	2,7	2,8	2,5	1,5	1,4	1,1	1,2	1,4	1,4
M-L-C- KD	49,4	39,4	38,1	38,9	38,5	40,2	38,7	37,7	37,3	36,1	36,0
S-V-MP	43,6	43,6	40	40	39,9	40,7	41,1	42,0	41,7	43,3	41,5
Block skillnad	-5,8	4,2	1,9	1,1	1,4	0,5	2,4	4,3	4,4	7,2	5,5

Om man väljer att jämföra med *valresultatet* 2018, där totalt 6 476 725 giltiga röster avgavs, vilka av partierna L, S, och SD har en statistiskt säkerställd *förändring* till februari 2019. Välj själv om du löser uppgiften med konfidensintervall eller med hypotesprövning. Använd $\alpha = 5\%$.

5. Du misstänker din (inte fullt så) goda vän Katarina för fusk Vid ett besök hos henne "råkar" du få med dig hennes favorittärning hem, och bestämmer dig för att testa den. Eftersom du verkligen inte vill anklaga någon för fusk ogrundat bestämmer du dig för att göra inte mindre än 100 tärningskast för att se hur många sexor det blir. På förhand bestämmer du dig för att kräva *mer än* 20 sexor för att våga påstå att hon är ohederlig.

Låt alltså testet vara att se om man får mer än 20 sexor vid 100 kast. Då går det att beräkna styrkefunktionen för testet, vilken ses i grafen på nästa sida. Du behöver inte beräkna styrkefunktionen, utan förväntas avläsa nödvändiga uppgifter ur grafen.

- Formulera nollhypotes, H_0 , och mothypotes (=alternativhypotes), H_a , för testet.
- Vilken är risken (=sannolikheten) att förkasta en sann nollhypotes (bestäm ur grafen)?
- Vilken styrka har testet för en tärning som har sannolikheten 20% att få en sexa (bestäm ur grafen)?
- Vilken är risken (=sannolikheten) för godtagandefel för en tärning som har sannolikheten 20% att få en sexa (bestäm ur grafen)?
- Vilken typ av fel, förkastelsefel eller godtagandefel, anser du vara värst att göra i det här fallet? Motivera.
- Till sist, efter alla förberedelser och överväganden ovan, så utför du testet. Efter 100 kast räknar du till 17 stycken sexor. Vilken slutsats kan du dra av detta resultat?



6. Vilket eller vilka påståenden är sanna? Vilket eller vilka är falska? Markera för varje påstående "S" om det är sant och "F" om det är falskt. Lämna blankt om du är osäker.
- (a) Korrelationskoefficienten, r , ligger alltid mellan noll och ett, dvs $0 \leq r \leq 1$ gäller alltid.
 - (b) Kvadraten på korrelationskoefficienten, r^2 , ligger alltid mellan noll och ett, dvs $0 \leq r^2 \leq 1$ gäller alltid.
 - (c) Ett dubbelsidigt konfidensintervall för genomsnittsnivå är alltid kortare än motsvarande dubbelsidiga prognosintervall (inom regressionanalys).
 - (d) Ett enkelsidigt konfidensintervall för genomsnittsnivå är alltid kortare än motsvarande enkelsidiga prognosintervall (inom regressionanalys).
 - (e) Regressionsparametern b_1 är ett mått på hur väl datapunkterna ligger längs en rät linje.
 - (f) Regressionsparametern b_1 i kvadrat, b_1^2 , är ett mått på hur väl datapunkterna ligger längs en rät linje.

Endast svar krävs i denna uppgift. Varje rätt svar ger en poäng och varje fel svar minus en poäng, dock kan totala poängsumman inte bli mindre än noll. Om du lämnar blankt blir det varken plus eller minus.