

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 24 mars 2022, kl. 14.00 – 18.00.

Kursens förväntade läranderesultat enligt kursplanen

Efter genomförd kurs ska 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 slumpvariabler, även med centrala gränsvärdessatsen, 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 (utan wifi-uppkoppling)

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

Frågor besvaras av Michael Hörnquist som besöker skrivsalen cirka kl. 15.00 och kl. 16.30. Svar och kortfattade lösningsförslag finns på Studieinfo senast kl. 20 på tentamensdagen. Skrivningsresultat meddelas senast femton arbetsdagar efter tentamenstillfället.

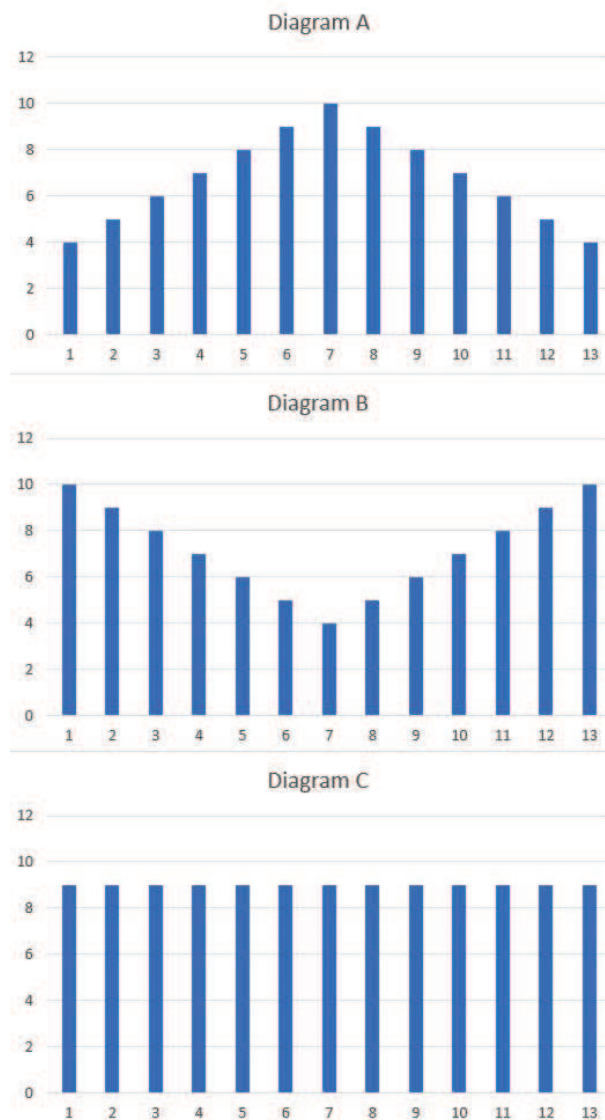
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs vid rättningen och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng, för betyg 3 ska dock minst 12 poäng komma från tentamen.

Svaren som lämnas in ska anges på bifogad svarsblankett, och poängsättningen kommer att utgå från att det verkligen står ett svar på den. För att säkerställa den egna prestationen ska dock den kalkyl som lett fram till givna svar bifogas svarsblanketten.

Lycka till!

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

1. Stolpdiagrammen nedan beskriver tre olika datamängder, A, B och C.



Mängderna har medelvärden μ_A , μ_B och μ_C samt standardavvikelser σ_A , σ_B och σ_C .

- Ordna medelvärdena i storleksordning, från minsta till största. Ange speciellt om några av värdena är lika stora.
- Ordna standardavvikelserna i storleksordning, från minsta till största. Ange speciellt om några av värdena är lika stora.

Avsikten är inte att du ska räkna ut något av dessa värden, utan resonera dig fram till svaren utifrån stolpdiagrammens utseende.

2. Låt oss anta att Sveriges befolkning består av tio miljoner människor, och att enligt folkbokföringen 53% av dem är män och 47% kvinnor (sedan finns de som inte identifierar sig med någondera könet, men den möjligheten finns ännu inte i officiell svensk statistik). Välj en person helt slumpmässigt ur Sveriges befolkning, och betrakta händelserna: A ="personen är man", B ="personen är kungen", C ="personen är färgblind".
 - (a) Bestäm sannolikheterna $\Pr(A|B)$ och $\Pr(B|A)$ samt förklara med ord vad de står för.
 - (b) Är händelserna A och B oberoende? Motivera svaret.
 - (c) Antag att 8% av männen och 0,7% av kvinnorna i Sverige är färgblinda. Bestäm $\Pr(C)$.

3. Längden på en graviditet kan approximativt beskrivas av en normalfördelad slumpvariabel med väntevärde $\mu = 266$ dagar och standardavvikelse $\sigma = 16$ dagar. En graviditet som avviker max 14 dagar från väntevärdet kallas att vara av *normal längd*. De tre vännerna Sigrid, Eva och Katarina samtalar och funderar tillsammans över sina graviditeter.
 - (a) Vad är sannolikheten att Sigrids graviditet ska bli av normal längd?
 - (b) Vad sannolikheten att alla tres graviditeter ska bli av normal längd?
 - (c) Vad sannolikheten att den genomsnittliga längden av de tre graviditeterna ska bli av normal längd?

4. En tillämpning av statistiska metoder är att undersöka om slumpprocesser verkligen är så slumpmässiga som de utger sig för. Här studerar vi kast med mynt, där utfallet i varje kast är antingen "krona" eller "klave". Tänk dig att du har kommit över ett specialmynt, och ska undersöka om det är välbalanserat så att sannolikheten för respektive utfall är 50%. Genomför ett försök med 100 kast, och räkna antalet "krona".
 - (a) Vad är förväntansvärdet för andelen "krona"?
 - (b) Hur stor är standardavvikelsen för andelen "krona"?
 - (c) Låt säga att försöket resulterar i 46 utfall "krona". Ange ett dubbelsidigt konfidensintervall för sannolikheten att få "krona" på konfidensnivån 95%.
 - (d) Hur många kast behöver göras för att konfidensintervallet ska bli hälften så brett?

Antag i (a) och (b) att myntet är välbalanserat, dvs att det är samma sannolikhet för "krona" som för "klave".

5. Hur stor är en "large" kopp kaffe? Den världsberömda kedjan Stjärnstålar, etablerad åtminstone över hela Östergötland, berömmar sig av att du som kund får 40 cl (centiliter) dryck när du beställer en "large" kopp. Din erfarenhet är att muggen i och för sig rymmer den volymen, men du upplever att den oftast är halvfull. Som statistikskunnig angriper du frågan vetenskapligt, och genomför en hypotesprövning för att försöka visa att mängden kaffe understiger 40 cl. Du besöker Stjärnstålar vid sju olika tillfällen, utrustad med ett måttglas, och finner att kaffemängden vid respektive tillfälle är 37, 41, 39, 36, 37, 42 cl.

Välj signifikansnivå 5%. Kan du baserat på dessa mätningar med hedern i behåll anklaga Stjärnstålar för fusk med kaffemängden?

Slutsatsen skriver du på svarsblanketten, medan själva hypotesprövningen skrivs på medföljande blad med beräkningar.

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) Om korrelationskoefficienten $r = 0,9$ innebär det att regressionslinjen lutar brantare än om korrelationskoefficienten $r = 0,4$.
 - (b) Om riktningskoefficienten $b_1 = 0,9$ innebär det att regressionslinjen lutar brantare än om riktningskoefficienten $b_1 = 0,4$.
 - (c) Om korrelationskoefficienten $r = 0$ innebär det att ingen regressionslinje kan bestämmas.
 - (d) p -värdet som åtföljer riktningskoefficienten i en regressionslinje är baserad på nollhypotesen $\beta_1 = 0$.
 - (e) Om alla punkter i ett spridningsdiagram ligger på en horisontell rät linje så är korrelationskoefficienten $r = 0$.
 - (f) Om alla punkter i ett spridningsdiagram ligger på en horisontell rät linje så är riktningskoefficienten $b_1 = 0$.

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.