

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 26 mars 2021, 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.

Frågor besvaras av Michael Hörnquist som finns tillgänglig på telefon 011 - 36 33 81 under hela skrivtiden. Svar och kortfattade lösningsförslag finns på Studieinfo efter att sista tentanden har lämnat in (troligen kl. 19.30). Skrivningsresultat meddelas senast femton arbetsdagar efter tentamenstillfället.

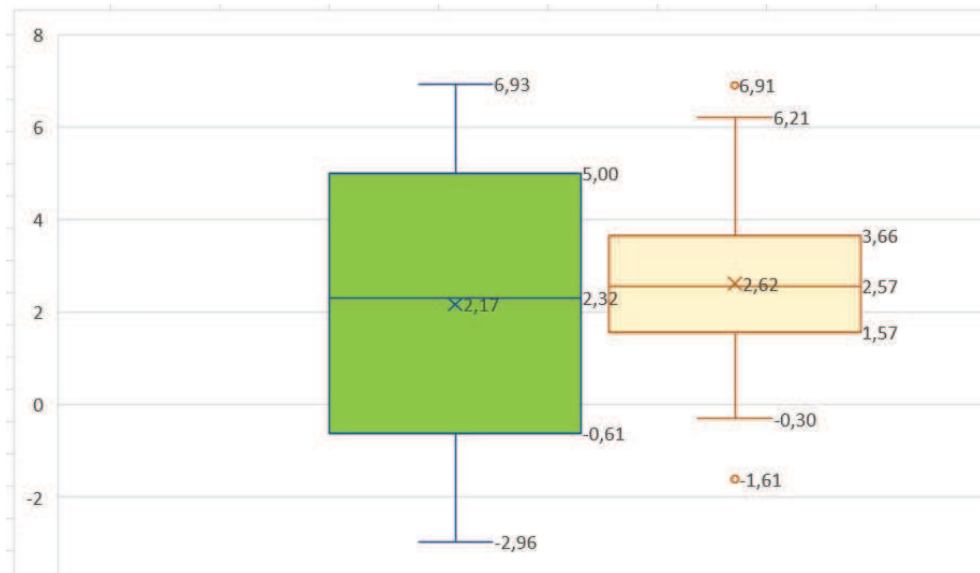
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs vid rättningen och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng.

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!

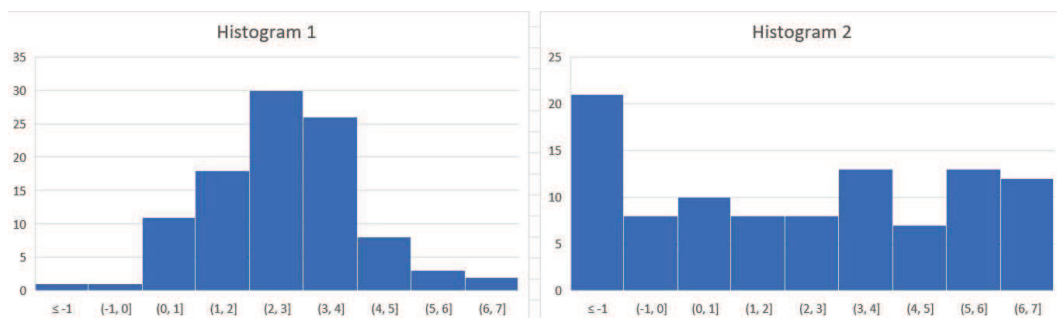
¹Kurslitteraturen "Tillämpad statistik – en grundkurs", Wahlin, Sanoma förlag, torde vara vanligast.

1. Vi har två datamängder, kallade A och B, var och en bestående av 100 tal. I figur 1 visualiseras dessa som lådagram (A till vänster och B till höger).



Figur 1: Lådagram för datamängderna A och B; A till vänster och B till höger.

- (a) Avläs ur figur 1 maxvärde, minvärde, median och medelvärde för respektive datamängd.
 (b) I figur 2 finns histogrammen för datamängderna. Vilket histogram hör till vilken datamängd? Motivering krävs.



Figur 2: Histogram för datamängderna A och B

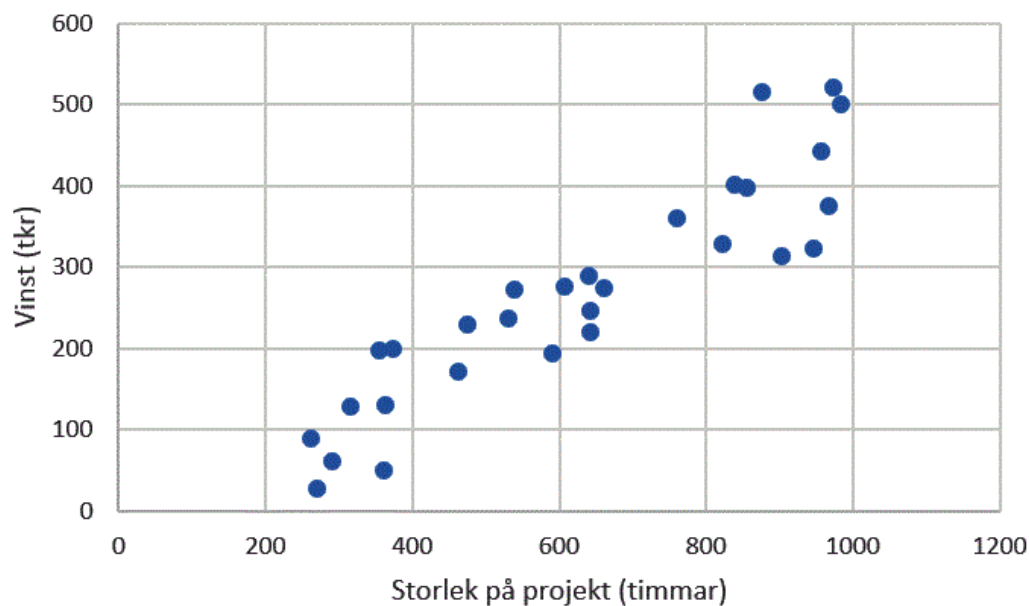
2. Din goda vän Margareta ger dig en fröpåse med två olika sorters rosfrön (låt kalla dem A och B), med olika sannolikhet för att gro. Frösört A gror med sannolikhet 70%, medan sannolikheten för frösört B att gro är okänd. Påsen innehåller 35% fröer av sort A och resten av sort B. Fröerna ser helt lika ut, det går inte att skilja mellan A och B. Du planterar många frön och ser att 80% av dem gror.

- (a) Hur stor är sannolikheten att ett frö av sorten B gror?
 (b) Du ser att det första fröet du planterade gror. Vad är sannolikheten att det var ett frö av sorten B?

3. Låt X och Y vara två oberoende exponentialfördelade slumpvariabler, $X \sim \exp(\lambda = 2, 0)$ och $Y \sim \exp(\lambda = 2, 5)$.
- (a) Bestäm sannolikheten att slumpvariabeln X tar ett värde som överstiger 8, dvs bestäm sannolikheten $\Pr(X > 8)$.
 - (b) Bestäm väntevärde och standardavvikelse för $Z = X + Y$, dvs för summan av slumpvariablerna.
 - (c) Vad kan sägas om fördelningen för slumpvariabeln $Z = X + Y$, är den exponentialfördelad, normalfördelad, eller något annat?
4. En hamburgare av typen "quarter pounder" bör rimligen väga ett kvarts pund. En internationell hamburgerkedja anger på sin webbsida att kötet i deras "quarter pounder" väger 113 gram. du bestämmer dig för att kontrollera uppgiften och ger dig ut till sju olika restauranger där du beställer en sådan hamburgare. Du väger respektive köttbit för sig, och får resultaten: 89 gram, 91 gram, 91 gram, 88 gram, 90 gram, 87 gram och 90 gram. Bestäm ett konfidensintervall på konfidensnivån 90% för medelvikten hos hamburgaren "quarter pounder" hos denna hamburgerkedja.
5. 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.
- Det så kallade p -värdet vid en hypotesprövning är:
- (a) sannolikheten att nollhypotesen är *falsk*.
 - (b) sannolikheten att nollhypotesen är *sann*.
 - (c) sannolikheten att godta en felaktig nollhypotes.
 - (d) sannolikheten att inte förkasta en falsk nollhypotes.
 - (e) ett värde som leder till förkastande av nollhypotesen om det understiger testets signifikansnivå α .
 - (f) ett värde som leder till att nollhypotesen anses bevisad om det understiger testets signifikansnivå α .

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.

6. Ett konsultföretag i bygg- och logistikbranchen har fått en förfrågan om ett uppdrag om 800 timmar. För att kunna göra en rimlig budget behöver man ett prognosintervall för förväntad vinst för ett sådant projekt, och ber som vanligt statistikavdelningen om en kalkyl.
- Där har man under en längre tid samlat in resultat för företagets olika uppdrag av högst varierande storlek, illustreras i spridningsdiagrammet nedan. Totalt sett har uppdragen i medel omfattat 630 timmar, men spridningen är stor, standardavvikelsen är 245 timmar.



(a) Bestäm ett 95% prognosintervall för förväntad vinst för det aktuella projektet om 800 timmar. Till din hjälp har du förutom spridningsdiagrammet även en regressionsanalys från Excel för samma data.

UTDATASAMMANFATTNING						
Regressionsstatistik						
Multipel-R	0,916654411					
R-kvadrat	0,840255309					
Justerad R-kvadrat	0,834338839					
Standardfel	55,54564131					
Observationer	29					
ANOVA						
	fg	KvS	MKv	F	p-värde för F	
Regression	1	438175,9798	438175,9798	142,019702	2,90726E-12	
Residual	27	83303,59326	3085,318269			
Totalt	28	521479,5731				
	Koefficienter	Standardfel	t-kvot	p-värde	Nedre 95%	Övre 95%
Konstant	-53,3200797	28,85027244	-1,84816555	0,0755603	-112,5159491	5,875789707
X-variabel 1	0,509946082	0,042790756	11,91720192	2,9073E-12	0,422146703	0,597745462

(b) För "X-variabel 1" anges att " p -värde" är 2,9073E-12 (dvs $2,9073 \cdot 10^{-12}$). Förklara vad det talet står för.