

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 7 juni 2022, kl. 8.00 – 12.00.

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. 9.00 och kl. 10.30. Svar och kortfattade lösningsförslag finns på Studieinfo senast kl. 14 på tentamensdagen. Skrivningsresultat meddelas senast femton arbetsdagar efter tentamenstillfället.

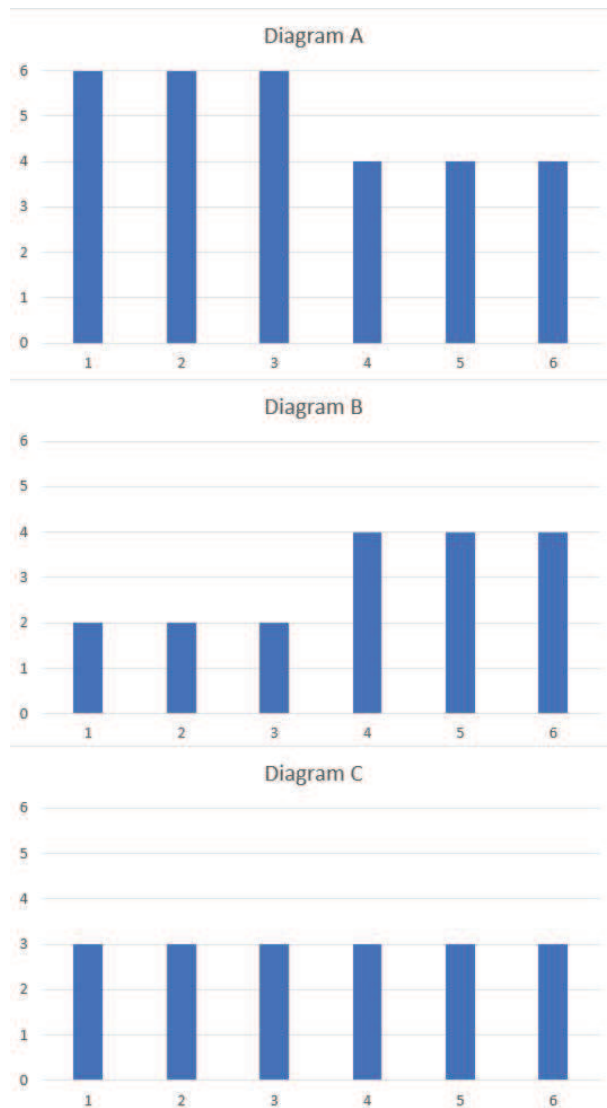
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs vid rättningen och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng, för betyg 3 ska dock minst 12 av dessa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

Lycka till!

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

1. Stolpdigrammen nedan beskriver tre olika datamängder, A , B och C , var och en bestående av heltal mellan 1 och 6.



Mängderna har medelvärden μ_A , μ_B och μ_C (dessa behöver inte räknas ut).

- Ordna medelvärdena i storleksordning, från minsta till största. Ange speciellt om några av värdena är lika stora.
- Beräkna standardavvikelsen för datamängd C (nu behöver du värdet på μ_C). Avgör själv om du bestämmer populationsstandardavvikelsen eller stickprovsstandardavvikelsen, men ange noggrant i svaret ditt val.

2. Vi återvänder till de tre vännerna från tentamen i mars 2022, Sigrid, Eva och Katarina, som funderar över sina graviditeter. Du behöver dock inte känna till vad de tidigare frågat sig för att kunna lösa denna uppgift.

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*.

- (a) Vad är sannolikheten att Eva ska föda för tidigt (dvs före normal längd)?
- (b) Vad är sannolikheten att minst en av de tre ska föda för tidigt?
- (c) Katarinas graviditet har redan varat i 260 dagar. Vad är sannolikheten att hennes graviditet kommer att bli av normal längd?

3. Låt X och Y vara två kontinuerliga slumpvariabler, $X \sim \text{likf}(-1, 1)$ ² och $Y \sim \exp(\lambda = 1)$.

- (a) Bestäm $\mu_X = E(X)$ och $\sigma_X = SD(X)$.
- (b) Bestäm $\mu_Y = E(Y)$ och $\sigma_Y = SD(Y)$.
- (c) Beräkna $\Pr(\mu_X - 2, 5\sigma_X \leq X \leq \mu_X + 2, 5\sigma_X)$.
- (d) Beräkna $\Pr(\mu_Y - 2, 5\sigma_Y \leq Y \leq \mu_Y + 2, 5\sigma_Y)$.

Kända samband för standardfördelningar kan med fördel användas.

4. Redan 24 februari 2022, samma dag som Ryssland i gryningen startade krig mot Ukraina, påbörjade opinionsinstitutet Novus en undersökning om det i Sverige fanns en opinion för att gå med i NATO. Man tillfrågade under två dagar 1003 personer, och av dessa ansåg 41% att landet borde gå med, 35% att vi inte borde gå med. Resterande var osäkra.

Bestäm två 99% konfidensintervall för andelen positiva respektive negativa till NATO-medlemskap för Sverige. Antag att Novus har gjort ett bra jobb och valt ut personer till undersökningen som ett OSU, obundet slumpmässigt urval.

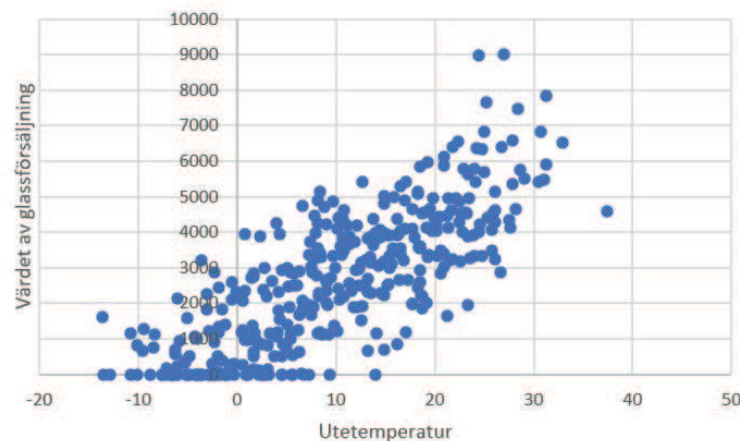
5. Vilket eller vilka påståenden är sanna? Vilket eller vilka är falska? Markera för varje påstående "S" om det är sant och "F" om det är falskt. Lämna blankt om du är osäker.

- (a) Det så kallade p -värdet anger sannolikheten för att nollhypotesen falsk.
- (b) Det så kallade p -värdet anger sannolikheten att få ett resultat som är åtminstone lika extremt som det man fick, givet att nollhypotesen är sann.
- (c) Det så kallade p -värdet anger hur stor andel av nollhypoteserna som inte kan förkastas, givet att hälften av dem är sanna.
- (d) Signifikansnivån, α , vid en hypotesprövning är sannolikheten att godta en felaktig nollhypotes.
- (e) Signifikansnivån, α , vid en hypotesprövning är sannolikheten att förkasta en sann nollhypotes.
- (f) Sannolikheten för godtagandefel kan beräknas ur styrkefunktionen.

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.

²Påminner om att beteckningen $\text{likf}(-1, 1)$ står för likformig fördelning, där täthetsfunktionen är nollskild endast på intervallet $[-1, 1]$.

6. Pontus vill göra om det "klassiska" experimentet att studera sambandet mellan temperatur och glassförsäljning. Han tar därför fram statistik från SMHI om högsta temperatur för varje dag under ett år, och jämför det med värdet av glassförsäljningen dag för dag under året på en Pressbyrå där han råkar känna ägaren. Pontus matar in värdena i Excel och ritar spridningsdiagrammet:



Sedan gör han en regressionsanalys. Utdatasammanfattningen blir:

UTDATASAMMANFATTNING						
Regressionsstatistik						
Multipel-R	0,73					
R-kvadrat	0,54					
Justerad R-kvadrat	0,54					
Standardfel	1321,19					
Observationer	365,00					
ANOVA						
	fg	KvS	Mkv	F	-värde för F	
Regression	1	736961080,9	7,37E+08	422,197	8,706E-63	
Residual	363	633630513	1745539			
Totalt	364	1370591594				
	Koefficienter	Standardfel	t-kvot	p-värde	Nedre 95%	Övre 95%
Konstant	1176,97	98,45	11,95	0,00	983,36	1370,58
X-variabel 1	140,61	6,84	20,55	0,00	127,15	154,07

Pontus vet att utdatasammanfattningen inte alltid ger all information som behövs för en regressionsanalys, så han tar också fram att medeltemperaturen är 9,5 grader och att temperaturens populationsstandardavvikelse är 10,4 grader.

- Vad är förklaringsgraden hos modellen?
- Vad är ekvationen för regressionslinjen?
- Ange en punktskattning för värdet av glassförsäljningen en högsommardag när temperaturen är 30 grader.
- Ange ett 90% konfidensintervall för värdet av försäljningen denna högsommardag med temperatur 30 grader.