

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 22 augusti 2020, 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.

Frågor besvaras av Michael Hörnquist som finns tillgänglig på telefon 011 - 36 33 81 under hela skrivtiden. Svar och kortfattade lösningsförslag finns på Studieinfo senast kl. 14.15. 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 förs in på bifogad svarsblankett (Urkund kommer att användas). Bakomliggande uträkningar skrivs för hand och skannas för att bifogas svarsblanketten så det är klart att det inte är gissningar som har gett svaren. Inlämning sker via Lisam, men om tekniken krånglar skickas de per epost till michael.hornquist@liu.se

Lycka till!

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

1. Vilket eller vilka påståenden är sanna? Vilket eller vilka är falska? *Kort motivering till varje svar krävs i denna uppgift, skrivs på svarsblanketten.*

- (a) Det går alltid att bestämma medianvärdet för en mängd av heltal.
- (b) Medianvärdet för en mängd av heltal är alltid ett heltal (om det går att bestämma).
- (c) Det går alltid att bestämma typvärdet för en mängd av heltal.
- (d) Typvärdet för en mängd av heltal är alltid ett heltal (om det går att bestämma).
- (e) Det går alltid att bestämma standardavvikelsen för en mängd av heltal (med minst tre element).
- (f) Standardavvikelsen för en mängd av minst tre heltal kan själv vara ett heltal.

OBS: Ändrade krav på redovisning mot tentamina mars 2020 och tidigare; inga minuspoäng för felaktiga svar. Motiveringen måste vara med egna ord, eller ett motexempel med siffror; det får inte vara direkt citat ur kurslitteraturen.

2. I samband med att jordens klimat blir allt varmare är det väsentligt att hålla ordning på vad som är effekter av klimatförändring och vad som är normala variationer. För ett bostadsområde i Halland vet man erfarenhetsmässigt att vart fjärde år inträffar så kraftiga regn att källarna blir vattenfyllda. Antag att regnoväder av den storleken inträffar enligt en poissonfördelning.

- (a) Vad är sannolikheten att det under ett enskilt år inte ska bli något så kraftigt regnoväder?
- (b) Vad är sannolikheten att det under enskilt år ska bli mer än ett sådant regnoväder?
- (c) Vad är sannolikheten att det under ett decennium (tio år) inte ska bli några sådana regnoväder alls?

3. En hiss i hus Täppan på campus Norrköping har följande display:



Antag att vikten hos en slumpmässigt vald medlem i Logistiksektionen är normalfördelad (välj själv rimligt väntevärde och standardavvikelse). På väg till årsmötet går 14 logistikstudenter in i hissen, utan en tanke på vare sig social eller fysisk distansering, eller för den delen att 14 är ett mer än 13. Jonas väljer trapporna, då han uppskattar att det är statistiskt säkert att den totala vikten överstiger 1000 kg. Beräkna sannolikheten att hissen blir överbelastad.

4. Tryckhållfastheten hos betong mäts ofta genom att kuber av standardmått utsätts för tryck tills de spricker sönder. En leverantör av betong med hållfasthetsklass C25/30 påstår att *mer än* 80% av hennes testkuber lever upp till kravet på en tryckhållfasthet om minst 30 MPa (megapascal, en miljon Newton per kvadratmeter).

Som stöd för sitt påstående bifogar hon laborationsprotokoll för tryckhållfastheten hos 40 olika betongkuber av klassen C25/30. Följande uppmätta tryckhållfastheter (enhet MPa) redovisas:

35,2	29,3	38,9	42,0	35,4	29,1	35,1	35,0	28,9	34,9
30,9	31,2	33,3	28,2	31,9	30,1	30,2	32,0	31,9	30,8
31,2	29,3	30,9	40,0	25,4	39,1	35,3	33,0	38,9	31,9
31,9	31,2	33,2	29,2	31,4	30,1	30,2	30,0	32,9	28,8

Undersök om mätprotokollet ger stöd för påståendet. Bilda ett lämpligt konfidensintervall för andelen kuber med en tryckhållfasthet överstigande 30 MPa. Välj själv konfidensnivå, och var noga med att ange ditt val. Formulera en tydlig slutsats om påståendet "*mer än* 80%".

5. Vi återvänder till kiosken Glädje-o-Lycka från tentamen i juni 2020.² Den påstår att fler vinner på deras Trisslotter än vad som kan förväntas (20%).
- (a) Konsumentverket vill kontrollera (igen) om det verkligen är sant, och gör en stickprovskontroll genom att köpa och skrapa 500 lotter. Hur många av dessa måste vara vinstlotter för att det ska anses statistiskt säkert (på signifikansnivån 5%) att kiosken Glädje-o-Lycka säljer fler vinstlotter än andra? Använd teorin för hypotesprövning för att besvara frågan.
- (b) Lotteriinspektionen vill fortfarande visa att påståendet är falskt, att det inte är fler än 20% som vinner på lotter köpta hos Glädje-o-Lycka. Förra tentamen gick de dessvärre med på ett test som var oanvändbart för dem (de försökte bevisa en nollhypotes). Nu funderar de på vad de kan göra istället. Deras önskan är att statistiskt bevisa att systemet med Trisslotter är säkert, och att andelen vinster hos Glädje-o-Lycka inte är större än för andra kiosker (20%). Ge dem ett gott råd för vad de kan göra.



KF - Eget arbete, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=656290>

²Du behöver inte känna till den för att lösa den här uppgiften.

6. Att sommarjobba med att ansvara för en glasskiosk och sälja glass på en badstrand är en utmaning. Tar du med för mycket glass en dag riskerar du att få kasta mycket i slutet av dagen, men tar du med för litet riskerar du att stå med arga kunder, dåligt rykte och bristande lönsamhet.

Studenten Katarina har dock klarat sin kurs i statistik, och bestämmer sig för att med statistiska modeller klara utmaningen. En vecka i början av juli, vecka 27, kontrollerar hon varje dag klockan 07:30 SMHI:s prognos för högsta temperatur den dagen samt lufttrycket på sin egen barometer. Innan hon packar ihop på kvällen kontrollerar hon hur mycket glass hon har sålt den dagen. Resultaten visas i tabell 1 och i figurerna 1 och 2 nedan.

Tabell 1: Temperatur, lufttryck och mängd såld glass för vecka 27

	Temp (°C)	Lufttryck (hPa)	Såld glass (kg)
söndag	15	955	22
måndag	25	1013	38
tisdag	25	1028	45
onsdag	29	1032	49
torsdag	30	1030	40
fredag	17	1004	35
lördag	23	1001	29

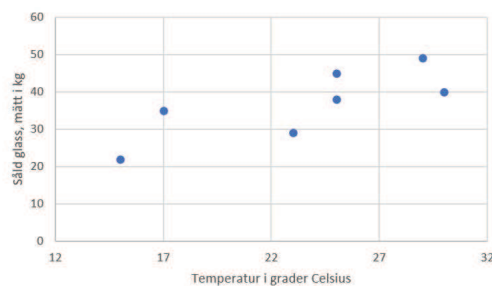


Figure 1: Glassförsäljning per dag vecka 27, temperatur som förklaringsvariabel

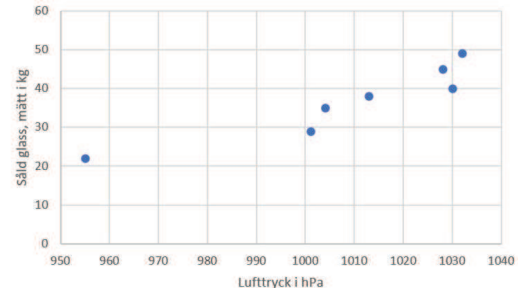


Figure 2: Glassförsäljning per dag vecka 27, lufttryck som förklaringsvariabel

För att kunna dra några slutsatser genomför hon två regressionsanalyser, en med temperatur som förklaringsvariabel och en med lufttryck. I båda fallen har hon mängd såld glass som responsvariabel. Resultaten presenteras i figurerna 3 och 4.

Regressionsstatistik						
Multipel-R	0,7778					
R-kvadrat	0,6050					
Justerad R-kvadrat	0,5261					
Standardfel	6,3524					
Observationer	7					
	Koefficienter	Standardfel	t-kvot	p-värde	Nedre 95%	Övre 95%
Konstant	7,1088	11,0136	0,6455	0,5471	-21,2026	35,4201
X-variabel 1	1,2697	0,4588	2,7676	0,0395	0,0904	2,4491

Figure 3: Regressionsanalys med Excel; temperatur förklaringsvariabel, mängd såld glass responsvariabel

Regressionsstatistik						
Multipel-R	0,9248					
R-kvadrat	0,8553					
Justerad R-kvadrat	0,8264					
Standardfel	3,8445					
Observationer	7					
	Koefficienter	Standardfel	t-kvot	p-värde	Nedre 95%	Övre 95%
Konstant	-282,8605	58,8185	-4,8090	0,0048	-434,0583	-131,6626
X-variabel 1	0,3169	0,0583	5,4373	0,0029	0,1671	0,4667

Figur 4: Regressionsanalys med Excel; lufttryck förklaringsvariabel, mängd såld glass responsvariabel

En dag i slutet av juli säger prognosen att en temperatur på 28 grader förväntas, och hon avläser lufttrycket till 1020 hPa.

- Hur mycket glass förväntar hon sig att kunna sälja denna julidag? Ett svar (punktskattning) per modell förväntas.
- Hur mycket glass bör hon ta med till kiosken denna dag, om hon vill kunna sälja så mycket som möjligt med bara 5% risk att glassen ska ta slut? Ett svar (konfidensintervall) per modell förväntas.
- Vilken av de två modellerna har högst förklaringsgrad? Motivering krävs.

Du kan vara betjänt av att veta att medeltemperaturen vecka 27 var 23,4 °C och stickprovsstandardavvikelsen 5,65 °C. För lufttrycket var motsvarande tal 1009,0 hPa respektive 26,9 hPa.