

Tentamen (TEN1) i **Statistik och sannolikhetslära** (TNIU66) 26 mars 2025, kl. 14.00 – 18.00.

Kursens förväntade läranderesultat enligt kursplanen

Efter genomförd kurs ska du kunna:

1. analysera och visualisera fördelningen hos en datamängd, såväl utan som med datorstöd.
2. beräkna sannolikheter för händelser med hjälp av begrepp, satser och definitioner som ingår i kursinnehållet.
3. genomföra statistisk analys med hjälp av begrepp, satser och definitioner som ingår i kursinnehållet, såväl utan som med datorstöd.

För betyg 3 på tentamen ska lärandemål 2 och 3 uppfyllas på grundläggande nivå och du ska använda begrepp, definitioner och satser som ingår enligt kursinnehållet. För betyg 5 ska lärandemål 2 och 3 uppfyllas på avancerad nivå och du ska med stor säkerhet använda begrepp, definitioner och satser som ingår enligt kursinnehållet. Avancerad nivå kan innebära att uppgiften innehåller flera delsteg som du själv behöver strukturera och begrepp från olika moment i kursen kan behöva kombineras. Stor säkerhet innebär att lösningar som presenteras ska vara välmotiverade och endast innehålla relevanta delar. Operativt översätts dessa kriterier till poänggränser.

Tentamen består av två delar, B Sannolikhetslära och C Statistik, med två respektive tre uppgifter per del. Varje uppgift ger 0 – 6 poäng. För betyget n krävs minst $5n$ poäng, varav minst 3 poäng från vardera delen B och C. Inget övrigt krav på fördelningen av poängen föreligger. Eventuell erhållen bonus från UPG1 påförs vid rättningen och ingår i den totala poängsumman.

Tillåtna hjälpmedel:

- Valfri bok inom statistik och sannolikhetslära¹
- Miniräknare av valfritt slag (utan wifi-uppkoppling)

Det får finnas anteckningar och markeringar i boken, inklusive ”pagemarkeringar” (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.

Lycka till!

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

Del B: Sannolikhetslära

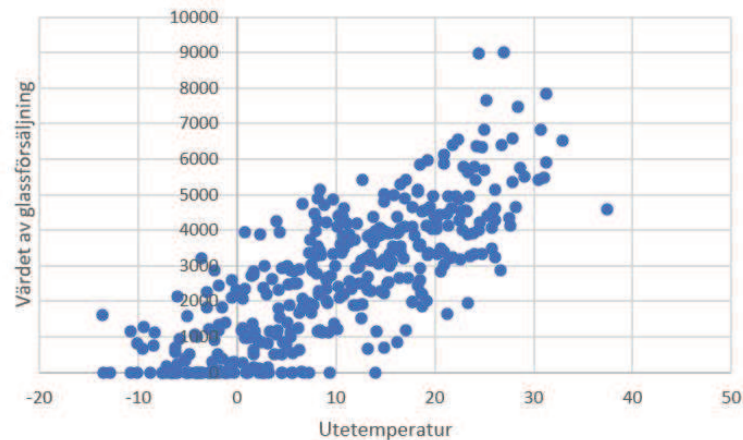
1. En vanlig kortlek består av 52 kort, jämt fördelade på fyra "färger", hjärter, spader, ruter och klöver. Det finns alltså 13 kort av varje "färg". Din gode vän Lukas blandar kortleken ordentligt så att korten ligger helt slumpmässigt i den, han lägger den sedan på bordet framför dig så att baksidorna på korten är uppåt.
 - (a) Vad är sannolikheten att översta kortet är en hjärter?
 - (b) Innan Lukas lade ned kortleken råkade du se att nedersta kortet är en spader. Vad är nu sannolikheten att översta kortet är en hjärter?
 - (c) Kortleken ligger kvar orörd framför dig, och du vet alltså att det nedersta kortet är en spader. Vad är sannolikheten att bland de *fem* översta korten finns *exakt två* som är hjärter?
2. Längder hos barn är en av många storheter som är approximativt normalfördelad. Betrakta nu en stor grupp fyraåringar som har medellängd 100 cm med en standardavvikelse om 15 cm.
 - (a) Vad är sannolikheten att ett barn i den gruppen som du slumpmässigt träffar på är längre än 110 cm?
 - (b) Välj ut fyra barn ur denna grupp, helt på måfå, i ett obundet slumpmässigt urval. Vad är sannolikheten att medellängden av dessa fyra barn är mer än 110 cm?

Del C: Statistik

3. Kommunen X-köping har en gymnasieskola där man alltid i samband med de allmänna valen ordnar skolval. Inför senaste valet gjorde de två eleverna Eiman och Ella inom ramen för sitt gymnasiearbete en stickprovsundersökning för att förutsäga resultatet och få lite fart på valarbetet. Skolan har 958 elever, och av dessa väljer Eiman och Ella ut 40 stycken helt slumpmässigt. Egendomligt nog väljer alla 40 att ställa upp i undersökningen. Det visar sig att åtta av de tillfrågade eleverna skulle rösta på Moderaterna, varvid Eiman och Ella med sina statistikkunskaper inser att punktskattningen för andel röstande på Moderaterna blir 20%.
 - (a) Hjälp Eiman och Ella att göra en mer användbar analys genom att bestämma ett dubbelsidigt konfidensintervall med konfidensnivån 90%.
 - (b) Förklara tydligt för Eiman och Ella vad konfidensintervall innebär, framför allt vad som är beroende av slumpvariabler.
 - (c) Eiman och Ella funderar nu på att fortsätta undersöka väljarsympatier för hela X-köping. Hur stort stickprov måste de då genomföra för att få samma felmarginaler som för skolundersökningen. Kommunen har 138 543 röstberättigade innevånare.
4. Eiman och Ella från uppgiften ovan (du behöver inte ha löst den för att kunna behandla den här uppgiften) bestämmer sig för att göra en opinionsundersökning för hela kommunen X-köping med 138 543 röstberättigade innevånare. Vad de främst vill se är om Moderaterna har *minskat* sedan förra valet när de fick 25% av rösterna, resultat från skolundersökningen i uppgift 3 tyder ju på det. För att ge ett seriöst intryck väljer Eiman och Ella ut 400 röstberättigade innevånare i X-köping i ett OSU. Av dessa uppger 92 att de skulle rösta på Moderaterna.

Hjälp Eiman och Ella genom att:

- (a) Geomföra ett hypotestest för att se om *nedgången* för Moderaterna sedan förra valet är statistiskt signifikant. Välj själv lämplig signifikansnivå.
- (b) Förklara tydligt för Eiman och Ella vilken slutsats som kan dras ur hypotestestet.
5. 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 med Excel 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						
	<i>fg</i>	<i>KvS</i>	<i>Mkv</i>	<i>F</i>	<i>-värde för F</i>	
Regression	1	736961080,9	7,37E+08	422,197	8,706E-63	
Residual	363	633630513	1745539			
Totalt	364	1370591594				
	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>	<i>Nedre 95%</i>	<i>Övre 95%</i>
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, medelförsäljningen 2323:-, temperaturens populationsstandardavvikelse 10,4 grader och försäljningens populationsstandardavvikelse 1732:-.

- (a) Hur stor är korrelationen mellan utetemperatur och värdet av försäljningen?
- (b) Ange ett 95% konfidensintervall för riktningskoefficienten.
- (c) Ange ett 90% konfidensintervall för medelvärdet av försäljningen högsommardagar med temperatur 25 grader.