

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 24 augusti 2019, kl. 8.00 – 12.00.

Kursens förväntade läranderesultat enligt kursplanen

Efter genomförd kurs skall 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 stokastiska variabler, även med centrala gränsvärdesatsen, 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

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

Frågor besvaras av Michael Hörnquist som besöker salen minst två gånger under skrivtiden, ca kl. 9 och 10.30. Svar och kortfattade lösningsförslag finns på Studieinfo efter skrivningens slut. Skrivningsresultat och visningstid meddelas via epost senast 15 arbetsdagar efter tentamenstillfället.

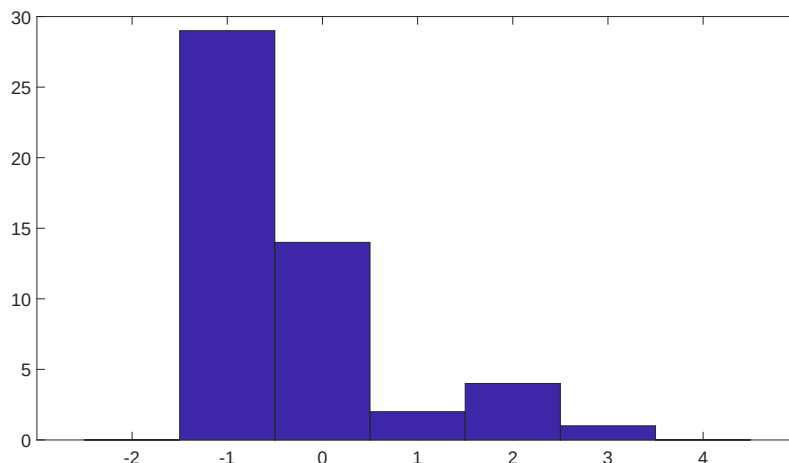
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs automatiskt och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng.

Redovisningen av uppgifterna skall vara sådan att det klart framgår vad du gör. Det innebär att förklarande text skall finnas med. Att enbart ge ett svar utan motivering, även om det är korrekt, är normalt inte tillräckligt för att uppgiften skall anses vara nöjaktigt löst. Å andra sidan behöver inte ett felaktigt svar, såvida det inte av någon orsak är orimligt, innebära att lösningen ej är godkänd. Väsentligt är att det bakomliggande resonemanget klart framgår samt är relevant.

Lycka till!

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

1. Nedanstående diagram visar fördelningen hos 50 stycken heltal mellan -1 och 3 . Bestäm medelvärde och median för talmängden.



Notera att y -axeln anger antal, och inte andelar. Staplarnas höjd är 29, 14, 2, 4 och 1.

2. Det anses känt (står i kurslitteraturen) att binomialfördelningen, $\text{bin}(n, \pi)$, kan approximeras med poissonfördelningen, $\text{poi}(\mu = n\pi)$ om vissa förutsättningar på n och π är uppfyllda. Enligt Wahlin är kravet $n > 20$ och $\pi < 0,05$, medan Løvås nöjer sig med $n > 10$ och $\pi < 0,1$. I den här uppgiften ska vi undersöka hur pass bra approximationen är i några specialfall.

Låt X och Y vara två slumpvariabler med fördelningarna $X \sim \text{bin}(n = 21, \pi = 0,04)$ och $Y \sim \text{poi}(\mu = 21 \cdot 0,04)$.

- Bestäm sannolikheterna $\Pr(X = k)$ för $k = 0$, $k = 3$ och $k = 25$.
 - Bestäm sannolikheterna $\Pr(Y = k)$ för $k = 0$, $k = 3$ och $k = 25$.
 - Bestäm de relativa felen i de tre fallen $k = 0$, $k = 3$ och $k = 25$. Med relativa felet menas kvoten (approximativt värde $-$ exakt värde)/(exakt värde); här ser vi värdet från binomialfördelningen som exakt och värdet från poissonfördelningen som approximativt, dvs vi söker $(\Pr(Y = k) - \Pr(X = k))/\Pr(X = k)$.
3. Ett företag i livsmedelsbranschen säljer ketchup på flaska. Enligt märkningen rymmer varje flaska $1/2$ liter, och en mer noggrann analys visar att vikten av ketchupen i en enskild flaska bäst beskrivs av en normalfördelad slumpvariabel med $\mu = 500$ ml och $\sigma = 15$ ml.
- Hur stor andel av flaskorna innehåller mindre än 490 ml ketchup?
 - Företaget vill ha färre missnöjda kunder, och beslutar sig för att fylla på med mer ketchup (öka μ). Målet är att 99% av flaskorna ska innehålla minst 500 ml ketchup. Vad behövs då för väntevärde för volymen? Anta att standardavvikelsen inte ändras.

4. I valet 2018 fick moderaterna 19,8% av rösterna. En opinionsundersökning genomförd i december samma år visade på att 18,8% stödde partiet. Gör två hypotesprövningar,
- (a) en för att undersöka om partiet har *ändrat* väljarstöd.
 - (b) och en annan för att undersöka om partiet har *minskat* väljarstöd.

Välj signifikansnivå $\alpha = 5\%$. Var noga med att ange vilka slutsatser du drar i respektive fall.

I riksdagsvalet 2018 deltog 6 535 271 personer (av 7 495 936 röstberättigade) och undersökningen från december 2018 besvarades av 5414 personer (av 10 742 tillfrågade). Bortfallsanalysen i opinionsundersökningen pekar på att bortfallet är åsiktsmässigt slumpmässigt.

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) p -värdet kan bestämmas ur styrkefunktionen.
 - (b) Vid hypotesprövning väljer man den hypotes som visar sig ha den högsta sannolikheten.
 - (c) Ett möjligt utfall vid ett hypotestest är att man (statistiskt) bevisar nollhypotesen.
 - (d) Syftet med ett hypotestest är att (statistiskt) bevisa att nollhypotesen är sann eller falsk.
 - (e) Ju högre värde på signifikansnivån (större α) vid ett hypotestest, desto sannolikare att få ett utfall så att nollhypotesen kan förkastas
 - (f) Ju högre värde på signifikansnivån (större α) vid ett hypotestest, desto sannolikare att få ett utfall så att alternativhypotesen blir (statistiskt) bevisad.

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. De två begreppen inflationstakt och arbetslöshet används ofta som indikatorer inom nationalekonomin för att beskriva vissa aspekter av samhället. För åren 1987 till 1992 (före "nittiotalskrisen") gällde i Sverige följande:

År	Inflationstakt (%)	Arbetslöshet (%)
1987	4,2	2,0
1988	5,8	1,8
1989	6,5	1,5
1990	10,4	1,6
1991	9,4	3,0
1992	1,8	5,3

Vi ska här använda dessa uppgifter för att göra förutsägelser med regressionsanalys. Låt inflationstakt vara förklaringsvariabel (x -variabel) och arbetslöshet responsvariabel (y -variabel). Regressionsanalys med Excel ger följande utdata:

UTDATASAMMANFATTNING						
Regressionsstatistik						
Multipel-R	0,57985					
R-kvadrat	0,33622					
Justerad R-kvadrat	0,17028					
Standardfel	1,32836					
Observationer	6,00000					
ANOVA						
	<i>fg</i>	<i>KvS</i>	<i>Mkv</i>	<i>F</i>	<i>p-värde för F</i>	
Regression	1	3,57516	3,57516	2,02611	0,22771	
Residual	4	7,05817	1,76454			
Totalt	5	10,63333				
	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>	<i>Nedre 95%</i>	<i>Övre 95%</i>
Konstant	4,20878	1,29598	3,24757	0,03145	0,61057	7,80699
X-variabel 1	-0,26385	0,18536	-1,42342	0,22771	-0,77850	0,25080

- Vad är förväntad arbetslöshet om inflationen hamnar på det mål riksbanken har om 2,0%?
- Ange ett 90% konfidensintervall för arbetslösheten i medel de år som inflationen är 2,0%.
- Enligt utdatasammanfattningen är korrelationskoefficienten (Multipel-R) positiv, närmare bestämt 0,57985. Det är högst egendomligt, då inflation och arbetslöshet brukar uppvisa ett negativt samband. Beräkna korrelationskoefficienten utifrån givna data med din miniräknare för att se om Excels beräkning verkligen kan stämma.