

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 20 augusti 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.

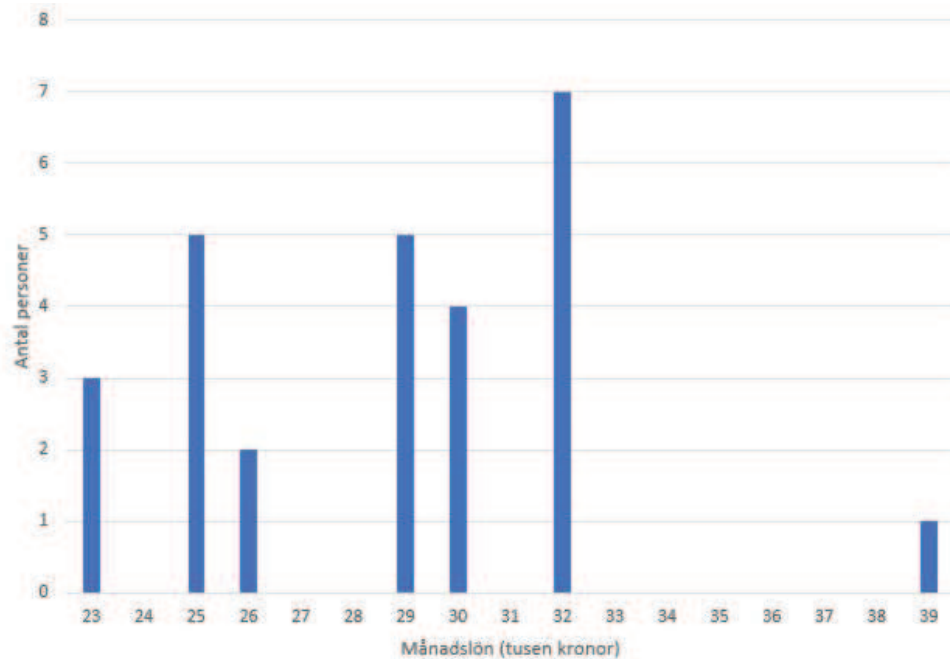
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs vid rättningen och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng, för betyg 3 ska dock minst 12 av dessa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

Lycka till!

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

1. Lönerna hos ett företag fördelar sig enligt:



Medelvärde är 28 667:- och standardavvikelsen är 3702:-.

Vid lönerevisionen överväger ledningen att antingen ge alla en höjning med 500:-, oavsett lönenivå, eller att höja alla löner med 1,7%.

- Ange median för lönerna och total lönekostnad för företaget *före* revisionen.
 - Ange medelvärde, median och standardavvikelse *efter* lönerevisionen om höjningen är 500:- för alla. Hur mycket blir företags totala lönekostnad i detta fall?
 - Ange medelvärde, median och standardavvikelse *efter* lönerevisionen om höjningen är 1,7% för alla. Hur mycket blir företags totala lönekostnad i detta fall?
2. I den här uppgiften studerar vi en vanlig sexsidig tärning, där sidorna som vanligt har mellan ett och sex "ögon". Låt X vara slumpvariabeln som anger antalet ögon på den sida som vid kast kommer upp på tärningen, med $X \sim \text{likf}(1, 6)$.

- Bestäm väntevärdet $\mu = E(X)$.
- Bestäm standardavvikelsen $\sigma = SD(X)$ (exakt svar eller svar med minst tre decimaler krävs).
- Bestäm sannolikheten $\Pr(X = \mu)$.
- Bestäm sannolikheten $\Pr(\mu - \sigma \leq X \leq \mu + \sigma)$.
- Bestäm sannolikheten $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$.

I (d) och (e) kan du använda $\sigma = 1,7$ om du inte har lyckats bestämma värdet i (b).

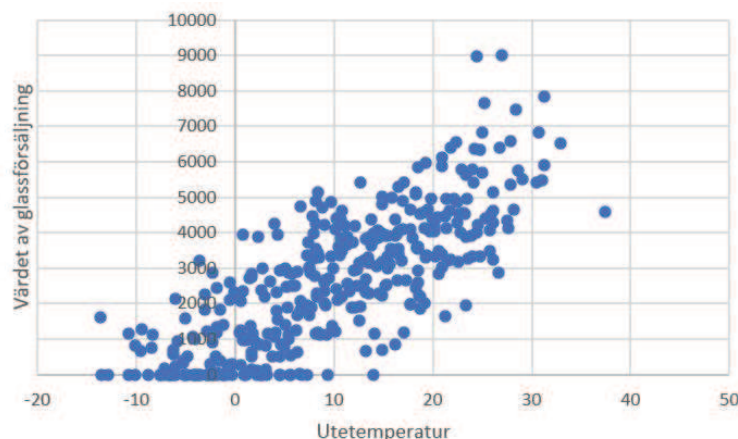
3. Den tid en spårvagn står vid en viss hållplats kan beskrivas med en normalfördelad slumpvariabel, $X \sim N(\mu = 1,50, \sigma = 0,45)$ minuter.
 - (a) Hur stor andel av spårvagnarna blir stående vid denna hållplats mer än två minuter?
 - (b) Vad är sannolikheten att medelvärdet av tiden vid hållplatsen för tre (oberoende) spårvagnar är mindre än en minut?

4. Antalet älgolyckor i Sverige *per år* under de nio åren 2010 – 2018 var i medeltal 5864 (källa: SLU). Bestäm ett dubbelsidigt 90% konfidenstervall för medelvärdet av årligt antal älgolyckor i Sverige.

5. Statistiska centralbyrån, SCB, genomför regelbundet partisympatiundersökningar. I maj 2022 gjordes den sista före riksdagsvalet i september och den visade bland annat på ett väljarstöd för Liberalerna (L) om 3,4% och för Kristdemokraterna (KD) om 5,2%.
 - (a) Genomför ett hypotestest för att se om L:s andel av väljarstödet är signifikant under riksdagssparren 4%. Slutsatsen skrivs på svarsblanketten medan själva testet bifogas på samma sätt som övriga motiveringar/beräkningar.
 - (b) Genomför ett hypotestest för att se om KD:s andel av väljarstödet är signifikant över riksdagssparren 4%. Slutsatsen skrivs på svarsblanketten medan själva testet bifogas på samma sätt som övriga motiveringar/beräkningar.

Genomför testen på signifikansnivån $\alpha = 0,1\%$ (vanligt är annars 5%, men vi vill vara extra säkra). Undersökningen baserar sig på ett obundet slumpmässigt urval om 9228 personer ur populationen folkbokförda i Sverige och röstberättigade i riksdagsval. Populationen omfattar 7,5 miljoner människor (källa: SCB).

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:



Du som skrev tentamen i juni 2022 känner igen inledningen av uppgiften, såhär långt är den identisk. Fortsättningen är dock helt annorlunda.

När Pontus gör en regressionsanalys med Excel slarvar han med datahanteringen och utdatasammanfattningen blir:

UTDATASAMMANFATTNING						
<i>Regressionsstatistik</i>						
Multipel-R	1					
R-kvadrat	1					
Justerad R-kvadrat	1					
Standardfel	3,77357E-13					
Observationer	365					
ANOVA						
	<i>fg</i>	<i>KvS</i>	<i>Mkv</i>	<i>F</i>	<i>-värde för F</i>	
Regression	1	1266800938	1,27E+09	8,9E+33	0	
Residual	363	5,16907E-23	1,42E-25			
Totalt	364	1266800938				
	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>	<i>Nedre 95%</i>	<i>Övre 95%</i>
Konstant	4,5475E-13	3,3072E-14	13,7504	6,35E-35	3,9E-13	5,198E-13
X-variabel 1	1	1,0602E-17	9,43E+16	0	1	1

- Vad är förklaringsgraden hos modellen (enligt utdatasammanfattningen)?
- Vad är ekvationen för regressionslinjen (enligt utdatasammanfattningen)?
- Hur märks att något är tokigt (flera möjliga svar finns här, välj ett)?
- Vilket slarv med data kan Pontus ha gjort?