Institutionen för teknik och naturvetenskap
Michael Hörnquist, 011-363381

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 23 mars 2023, kl. 14.00 – 18.00.

Kursens förväntade läranderesultat enligt kursplanen
Efter genomförd kurs ska du kunna:

1. analysera och visualisera fördelningen hos en datamängd.
2. beräkna sannolikheter för vissa händelser med hjälp av teoretiska begrepp som ingår i kursinnehållet.
3. beräkna punktskattningar och konfidensintervall.
4. genomföra hypotesprövning
5. genomföra enkel, linjär regressionsanalys.
6. använda datorstöd för beräkningar där det är relevant.

Tillåtna hjälpmedel:

- Valfri bok inom statistik och sannolikhetslära[1]
- 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. 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.

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 poäng av dessa komma från tentamen. Inget övrigt krav på fördelningen av poängen föreligger. Svaren som lämnas in ska anges på bifogad svarsblankett, och poängsättningen kommer att utgå från att det verkligen står ett svar på den. Den kalkyl som lett fram till givna svar bifogas svarsblanketten.

Lycka till!

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

1. Det får anses välkänt att populationsstandardavvikelsen, σ, för en mängd tal $\{x_1, x_2, \cdots, x_N\}$ definieras som

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2},$$

där μ är det aritmetiska (dvs vanliga) medelvärdet av talen i mängden. Kvadratroten brukar motiveras med att man vill att standardavvikelsen ska ha samma enhet som elementen i mängden, till exempel om varje x_i anger en längd mätt i meter, så vill man att spridningsmåttet också ska ha enheten meter (utan kvadratroten skulle det bli kvadratmeter, eftersom varje term kvadreras).

   Det här kan tyckas egendomligt. Först kvadrera och sedan roten ur. Låt oss betrakta en alternativ standardavvikelse, σ_a, där vi avstår från såväl kvadreringen som kvadratroten. Denna nya "standardavvikelse" ges då av formeln

$$\sigma_a = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu).$$

   Visa att det finns minst en allvarlig nackdel med denna alternativa definition genom att lösa uppgiferna nedan:

   (a) Beräkna σ och σ_a för mängderna $\{1, 2, 3\}$ och $\{0, 1, 2, 5\}$.

   (b) Beräkna σ_a för mängden $\{x_1, x_2, \cdots, x_N\}$.

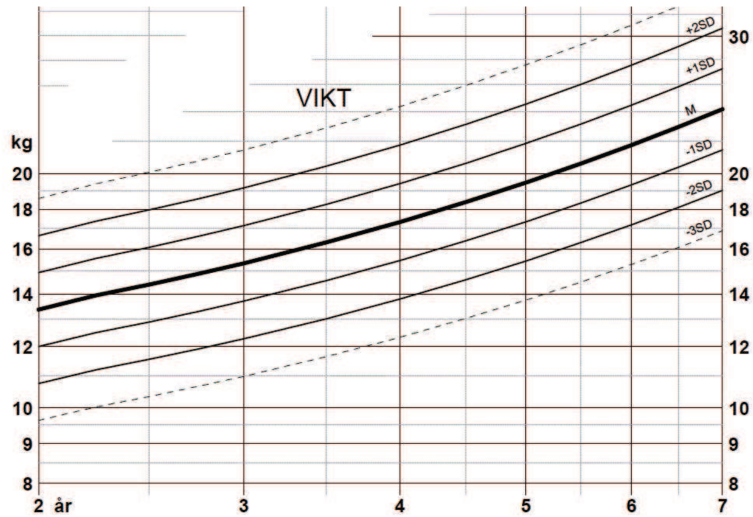
2. Den nyöppnade hamburgerrestaurangen Mama's gästas av de tre vännerna Margareta, Eva och Thomas. De har inget lätt val framför sig, då man här får välja mellan tre olika bröd, åtta olika grönsaker och fem olika protein. I en måltid ingår ett bröd, ett valfritt antal grönsaker och ett protein. Margareta önskar enbart en grönsak, utöver bröd och protein, medan Eva och Thomas önskar två grönsaker var, samt bröd och protein. Bestäm:

   (a) Antal olika kombinationer Margareta kan välja.

   (b) Antal olika kombinationer Eva kan välja. Hon väljer alltid två olika grönsaker, men bekymrar sig inte om vilken som läggs på först.

   (c) Antal olika kombinationer Thomas kan välja. Han väljer alltid två olika grönsaker, och anser att "gurka och tomat" är något helt annat än "tomat och gurka", dvs annan ordning på grönsakerna innebär en ny kombination.

3. Vikten för pojkar i Sverige i åldrarna två till sju år följer nedanstående diagram.[2]



Här fokuserar vi på tvååringarna.

(a) Ange medelvikt och standardavvikelse för tvååriga pojkar genom att läsa ut detta ur diagrammet. Eftersom det inte går att se exakta värden i ett diagram räcker det med ungefärliga värden.

(b) Hur stor andel av de tvååriga pojkarna väger mindre än 16 kg?

(c) Om vi vill studera den grupp av pojkarna som väger mest, den översta procenten, vilken vikt svarar det mot att de måste väga mer än?

4. 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) Om man vill höja konfidensnivån utan att ändra på något i sin stickprovsundersökning kan det åstadkommas genom att göra konfidensintervallet längre.

(b) Om man vill höja konfidensnivån utan att ändra på något i sin stickprovsundersökning kan det åstadkommas genom att göra konfidensintervallet kortare.

(c) Student t-fördelningen övergår i normalfördelningen $N(\mu = 0, \sigma = 1)$ när antalet frihetsgrader växer mot oändligheten.

(d) En fördubblad konfidensnivå ger en fördubblad längd på konfidensintervallet.

(e) En fördubblad signifikansnivå ger en fördubblad längd på konfidensintervallet.

(f) Vid stickprovsundersökningar av andelar ger ett dubbelt så stort stickprov ett konfidensintervall som är fyra gånger smalare (antag dubbelsidigt intervall).

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.

[2] Återgiven med tillstånd från PC PAL, www.tillvaxtkurvor.se.

5. Det inte så välkända läkemedelsföretaget ”Sundhet Och Friskhet För Alla”, SOFFA, har utvecklat en ny huvudvärksmedicin som de önskar testa. Bästa medicinen som säljs idag botar 60% av patienterna, så för att kunna sälja den nya medicinen behöver SOFFA kunna statistiskt bevisa att deras nya medicin är bättre än så. De låter därför utföra ett test enligt noggrant utprövad erfarenhet, och finner att av 228 deltagare i testet blev 160 av med sin huvudvärk.

   Genomför ett hypotestest på nivån $\alpha = 5\%$ för att se om angivet resultat är tillräckligt för att SOFFA ska kunna hävda att deras medicin är den bästa på marknaden. Förklara tydligt vilken slutsats du drar av hypotestestet.

6. En form av cement baseras på mineralet trikalciumsilikat, som bland annat ger snabb hållfasthetstillväxt. När cement stelnar utvecklas värme, hur mycket anses bero bland annat på mängden av det mineralet. För att närmare utreda detta utförs en försöksserie med tolv oberoende experiment där man lät andelen trikalciumsilikat variera mellan 27% och 70%. Resultatet blev följande mätserie:

x_i:	0,27	0,30	0,32	0,32	0,40	0,47	0,51	0,55	0,56	0,56	0,67	0,70
y_i:	18,8	17,8	21,0	17,3	20,0	27,7	22,9	22,3	26,1	25,0	27,1	24,6

   Den övre raden anger andelen trikalciumsilikat och den undre mängden utvecklad värme (angett i Joule per gram cement). Som ett första steg i analysen beräknas medelvärden och standardavvikelse för x- respektive y-värden, med resultaten $\bar{x} = 0,469$ och $s_x = 0,146$ samt $\bar{y} = 22,55$ J/gram och $s_y = 3,609$ J/gram. En linjär regressionsmodell ansätts, och ur Excel erhålls:

UTDATASAMMANFATTNING						
Regressionsstatistik						
Multipel-R	0,802					
R-kvadrat	0,644					
Justerad R-kvadrat	0,608					
Standardfel	2,259					
Observationer	12					
	Koefficienter	*Standardfel*	*t-kvot*	*p-värde*	*Nedre 95%*	*Övre 95%*
Konstant	13,3	2,3	5,8	1,66E-04	8,2	18,3
X-variabel 1	19,8	4,7	4,3	1,69E-03	9,4	30,1

   (a) Vilket väsentligt moment att göra innan man gör en regressionsanalys har vi hoppat över?

   (b) Hur väl korrelerar andelen trikalciumsilikat med uppmätt värmeutveckling, dvs vad är korrelationskoefficienten?

   (c) Framgent planeras att alltid använda en tillblandning av cement med 50% trikalciumsilikat. Ange ett 95% konfidensintervall för hur stor värmeutveckling man då kan förvänta sig i medeltal.