



## Försättsblad till skriftlig tentamen vid Linköpings Universitet

<b>Datum för tentamen</b>	2011-08-17
<b>Sal (1)</b> Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER3
<b>Tid</b>	8-12
<b>Kurskod</b>	732G81
<b>Provkod</b>	TENT
<b>Kursnamn/benämning</b> <b>Provnamn/benämning</b>	Statistik Tentamen
<b>Institution</b>	IDA
<b>Antal uppgifter som ingår i tentamen</b>	6
<b>Jour/Kursansvarig</b> Ange vem som besöker salen	Karl Wahlin
<b>Telefon under skrivtiden</b>	0709-719096
<b>Besöker salen ca kl.</b>	
<b>Kursadministratör/kontaktperson</b> (namn + tfnr + mailaddress)	carita.lilja@liu.se 1463
<b>Tillåtna hjälpmedel</b>	Valfri räknedosa, formelsamling utan anteckningar
<b>Övrigt</b>	
<b>Vilken typ av papper ska användas, rutigt eller linjerat</b>	Rutigt
<b>Antal exemplar i påsen</b>	6



# Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

---

Kurskod och namn:	732G81 Statistik för internationella ekonomer
Datum och tid:	2011-08-17, 8-12
Jourhavande lärare:	Kalle Wahlin
Tillåtna hjälpmedel:	Valfri räknedosa, formelsamling utan anteckningar.
Betygsgränser:	Tentamen omfattar totalt 40p. Godkänt från 24p, väl godkänt från 32p. Siffrorna i uppgifterna är påhittade.

---

Redovisa och motivera tydligt alla dina lösningar!

## Uppgift 1 (5p)

Ett företag som tillverkar motorgräsklippare (av den typ som man går med) hävdar att i snitt så fungerar maskinerna 2000 timmar och att standardavvikelsen är 250 timmar. Man säger också att andelen maskiner som fungerar mer än 2200 timmar är större än 30 %. För att undersöka detta tar man och testkör 10 gräsklippare. Här är resultatet ordnat från den kortaste till den längsta funktionstiden. Det är rimligt att anta att funktionstiden är normalfördelad.

Funktionstider:

1571 1574 1798 1890 1968 1982 2050 2108 2154 2793

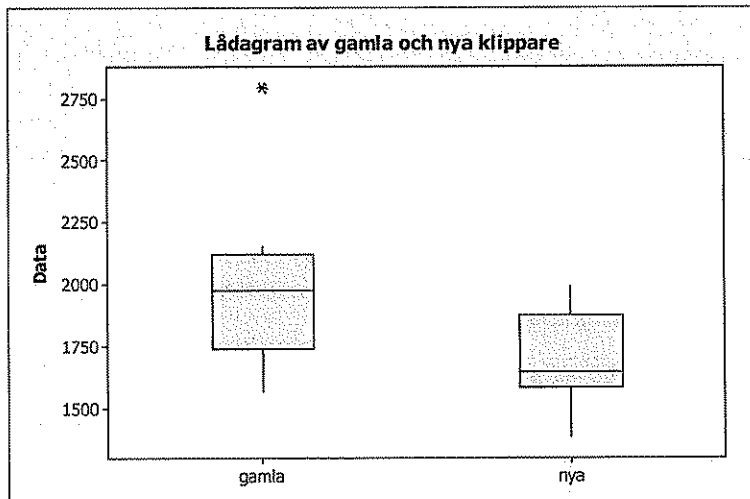
- Beräkna medelvärde, median samt standardavvikelse för de 10 gräsklipparna som testats. Vilket lägesmått är bäst för dessa data, medelvärdet eller median? Förklara. (3p)
- Beräkna sannolikheten att en slumpmässigt vald gräsklippare fungerar längre än 2200 timmar under förutsättning att uppgifterna om funktionstid som företaget ger stämmer. Verkar påståendet om 30% rimligt? (2p)

## Uppgift 2 (6p)

Gräsklipparna är för dyra att tillverka om de ska ha så lång förväntad funktionstid, så man försöker därför tillverka billigare klippare med kortare funktionstid. 10 av den nya typen av klippare testkördes. Gräsklipparna från uppgift 1 kallas 'gamla' och från uppgift 2 'nya'. Använd alltså data från både uppgift 1 och 2.

Funktionstider nya klippare:

1993 1744 1937 1658 1497 1858 1636 1387 1618 1615



- a) Förklara hur lådagrammen är uppbyggda i grafen ovan. Kan man med hjälp av grafen ana att det finns skillnad mellan maskinerna? Motivera. (2p)
- b) Pröva på 10% signifikansnivå om de nya gräsklipparna har kortare funktionstid jämfört med de gamlas förväntade funktionstid. Ställ upp hypoteser, genomför hypotesprövningen och dra slutsatser med ord. (3p)
- c) Förklara med ord vad  $p$ -värde är och hur det används. (1p)

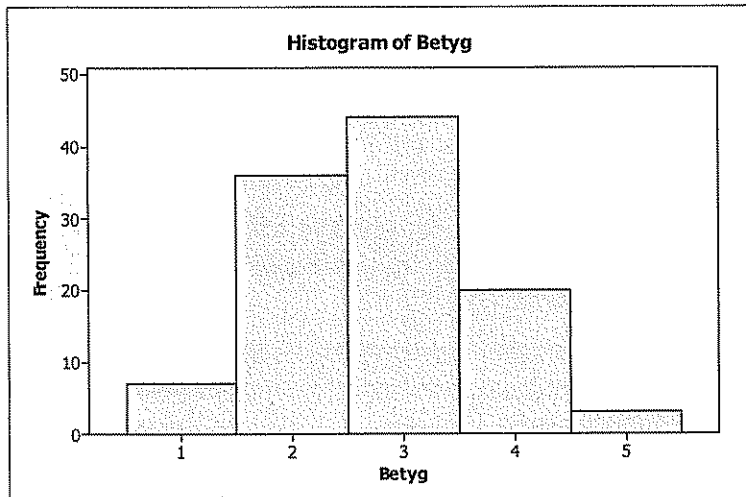
### Uppgift 3 (5p)

På företaget som tillverkar gräsklippare ville man undersöka hur stor andel av de anställda som tycker att det satsas tillräckligt mycket på friskvård. Man valde slumpmässigt ut 110 anställda. Av 65 män svarade 42 ja och bland de 45 kvinnorna svarade 18 ja.

- a) Bilda ett 95% konfidensintervall för skillnaden mellan andelen män och kvinnor som tycker att det satsas tillräckligt på friskvård. Tolka resultatet. (2p)

Man bad de 110 anställda att betygsätta fräschheten på duschutrymmet för de anställda. Betyg 1 är mycket ofräscht, upp till betyg 5 som är mycket fräscht. Låt  $\mu$  beteckna alla anställdas medelbetyg. Om det visar sig att medelbetyget ligger under 3 så måste förstås något göras för att få det fräschare. Antag att det totala antalet anställda är mycket stort.

Först ett histogram av resultatet:



Medelbetyg blev 2,7818 och standardavvikelsen blev 0,9125 för de 110 svaren.

Hypoteser:

$$H_0 : \mu \geq 3$$

$$H_a : \mu < 3$$

b) För att kunna pröva hypoteserna ovan måste centrala gränsvärdessatsen användas. Förklara varför den kan användas i detta fall. (1p)

c) Pröva hypoteserna ovan på 1 % signifikansnivå. Ställ upp hypoteser, genomför hypotesprövningen och dra slutsatser med ord. (2p)

#### Uppgift 4 (5p)

Ett företag arrangerar delfinsafari med båt. Man hävdar att sannolikheten är 60% att man får se delfiner på varje tur, och varje år arrangeras ett mycket stort antal turer.

- a) Vi väljer slumpmässigt 10 turer. Vad är sannolikheten för att man får se delfiner på fler än 8 av dessa? (2p)

Ibland får man också se valar. Sannolikheten för detta är enligt företaget 20%. Det förekommer att man på samma tur får se både valar och delfiner, men det råder oberoende mellan händelserna att se valar respektive delfiner.

- b) Åskådliggör händelserna "se valar under en tur" respektive "se delfiner under en tur" i ett Venndiagram. Är händelserna disjunkta? (1p)
- c) Vad är sannolikheten för att man på en slumpmässigt vald tur får se både valar och delfiner? (1p)
- d) Vad är sannolikheten för att man på en slumpmässigt vald tur får se antingen valar eller delfiner? (1p)

## Uppgift 5 (5p)

Skattemyndigheten i en kommun vill undersöka om det finns några skillnader mellan egenföretagare och anställda (ej egenföretagare) i fråga om man fuskat med deklARATIONEN. Man väljer slumpmässigt ut 89 egenföretagares och 123 anställdas deklARATIONER och undersöker dessa extra noga. 11 av egenföretagarna och 16 av de anställda har brister i deklARATIONEN.

- Ställ upp en korstabell uttryckt i absoluta frekvenser där raderna ska vara huruvida man är egenföretagare eller anställd och kolumnerna huruvida man fuskat med deklARATIONEN eller ej. Tabellen ska även innehålla marginalfrekvenserna. (2p)
- Undersök om det finns skillnader mellan egenföretagare och anställda i fråga om man fuskat med deklARATIONEN genom ett chitvåtest. Ställ upp hypoteser, genomför hypotesprövningen och dra slutsatser med ord. Använd 5% signifikansnivå. (3p)

## Uppgift 6 (14p)

Regressionsanalys används ofta för att studera utvecklingen över tiden för en variabel. Det inses att begreppen population och stickprov inte låter sig definieras lika enkelt i sådana fall, men trots det använder vi regressionsanalys på samma sätt.

Vi studerar här en fonds utveckling mellan år 2000 och 2009, mätt som en fondandels värde i kronor den sista december varje år.

År	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Fondvärde (kr)	105	121	132	119	147	126	151	164	188	162

- Åskådliggör data i ett spridningsdiagram. (2p)
- Beräkna korrelationskoefficienten mellan fondvärde och tid. Välj som  $y$  en fondandels värde och som förklaringsvariabel  $t$  ett tidsindex satt till 1 för år 2000. (2p)
- Beräkna  $a$  och  $b$  i en enkel linjär regressionsmodell. (3p)
- Tolka  $a$  och  $b$  med ord. (1p)
- Är  $\beta$  signifikant skild från 0? Använd 5% signifikansnivå. (3p)
- Beräkna modellens förklaringsgrad. Är modellen välanpassad? (1p)
- Beräkna ett lämpligt 95% intervall för det förväntade värdet på en fondandel år 2010. (2p)