

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I REGRESSIONSANALYS OCH UNDERSÖKNINGSMETODIK

DELKURS 1, REGRESSIONSANALYS OCH TIDSSERIEANALYS

2021-06-02

Skrivtid: 09.00-14.00. **Inlämningstid:** 14.00-15.00

Godkända hjälpmedel: Miniräknare, dator, kurslitteratur, föreläsninganteckningar och språklexikon, formelsamling och statistiska tabeller (bifogas).

Obs! Det är inte tillåtet att ta hjälp av andra personer under skrivningen

Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.

Resultatet meddelas senast den 16 juni.

Lösningförslag till tentamensuppgifterna läggs ut på Athena strax efter tentamen

Kontakt med examinator under tentamen: För eventuella frågor om innehållet i tentan kan du kontakta examinator under pågående tentamen på mail: Hans.Nyquist@stat.su.se. Inkommande mailfrågor besvaras kontinuerligt under tentans gång. Om examinator behöver informera om någonting under tentan görs detta till din registrerade mailadress. Kontrollera därför din mail under tentans gång.

Observera att praktisk hjälp endast finns tillgänglig under tentans första timme på mailadressen expedition@stat.su.se. Läs noggrant bifogad instruktion för inlämning av tentan. Där finns all nödvändig information om inlämning, anonymkod etc. Om du trots instruktionerna skulle få problem att lämna in tentan, maila istället tentan till tenta@stat.su.se. Detta görs dock bara i undantagsfall.

Uppgift 1. (25 poäng)

Ett företag undersöker vinsten som funktion av produktionsvolym. Under 12 veckor observerade man därför antal producerade enheter, x , och intäkter, y (i tusentals kr). Observationerna framgår av nedanstående tabell

Antal producerade enheter	Intäkter (tkr)
x	y
5	25
5	35
8	70
10	96
10	108
12	106
15	125
15	135
15	120
18	106
20	90
20	95

- Gör en lämplig figur över observationerna och beräkna urvalskovariansen och urvalskorrelationen mellan variablerna.
- Sätt upp en regressionsmodell med intäkter som beroende variabel och producerade enheter som förklaringsvariabel. Ange fullständiga antaganden. Ge en tolkning av modellens parametrar. Är sambandet kausalt? Motivera!
- Bestäm minstakvadratskattningarna av modellens parametrar och R^2 . Tolka det erhållna värdet på R^2 .
- Testa hypotesen att lutningsparametern är 0. Använd signifikansnivån 5 procent.
- Utvärdera modellen genom att analysera lämpliga residualplottar.

Uppgift 2. (25 poäng)

Sambandet mellan efterfrågan (y) och pris (x) på en elektronikprodukt analyserades med hjälp av en enkel linjär regressionsmodell. Med $n = 14$ observationer erhöles totala variationen av beroendevariabeln till $SS_T = 83$ och residualkvadratsumman till $SS_E = 18$. Lutningsparametern β uppskattades till ett negativt värde.

- a) Bestäm residualvariansen, s_e^2 .
- b) Bestäm förklaringsgraden R^2
- c) Pröva hypotesen $H_0 : \beta = 0$ mot $H_A : \beta < 0$ med ett t -test. Använd signifikansnivån 5 procent.

Uppgift 3. (15 poäng)

Förklara följande begrepp med egna ord. Förklaringarna ska vara fullständiga men kortfattade. Rita figur vid behov.

- a) Heteroskedasticitet
- b) F -test vid enkel och multipel linjär regression
- c) Konfidensintervall
- d) Löpande medelvärde
- e) Säsongsrensning

Uppgift 4 (20 poäng)

Företaget Statistiktjänst har varit börsnoterat på Nasdaq i New York under en tid. Under en följd av 12 börsdagar noterades företagens aktier för följande belopp (dollar):

Dag	1	2	3	4	5	6	7	8	9	10	11	12
Värde	5,00	4,98	5,00	5,00	4,93	4,98	4,97	4,92	4,86	4,87	4,91	4,89

- a) Sätt upp en AR(1) modell och skatta dess parametrar (tips: använd minst fyra decimaler vid beräkningarna).
- b) Gör prognoser för aktiekursen en respektive två börsdagar framåt i tiden (y_{13} och y_{14}).
- c) Tidsserien varierar kring ett visst tal och prognoserna närmar sig det talet. Detta tal är tidsseriens ekvilibrium. Bestäm en uppskattning av ekvilibrium för aktiekursen.

Uppgift 5. (15 poäng)

- a) Förklara med egna ord begreppet logodds samt beskriv hur logodds beror av en eller flera förklaringsvariabler i en logistisk regressionsmodell.
- b) I en undersökning av hur antalet studietimmar per vecka i en kurs i statistik påverkar resultatet i den avslutande tentamen användes logistisk regression. I modellen är antal studietimmar per vecka förklaringsvariabel. Beroendevariabeln är en Bernoullivariabel som indikerar tentamensresultatet (1 för godkänd och 0 för underkänd). Följande resultat erhöles vid estimation av modellen:

Analysis of Maximum Likelihood Estimates					
Coefficient	DF	Estimate	SE Coef	z-value	P-Value
Constant	1	-2,1973	0,1610	-13,65	0,000
Studietimmar	1	0,1792	0,0182	9,85	0,000

Bestäm logodds för händelsen 'godkänd' vid studietiden 10 timmar per vecka respektive vid 20 timmar per vecka.

- c) Bestäm sannolikheten att bli godkänd vid 10 respektive vid 20 studietimmar per vecka.
- d) Har förklaringsvariabeln studietid en signifikant påverkan på sannolikheten att bli godkänd?