STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)
Bo Wallentin

VT2017

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2017-04-20

Skrivtid:	kl. 16.00 - 21.00
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Bifogade hjälpmedel:	Häftet *Formelsamling och Tabeller över statistiska fördelningar* (återlämnas efter skrivningen)

- Tentamen består av 7 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.
- **Uppgift 1 – 5**: Svar lämnas på särskild **SVARSBILAGA,**
  - Flervalsfrågor där ett av fem alternativ är korrekt svar.
  - Har fler än ett svarsalternativ markerats för en deluppgift ges noll poäng.
  - Uträkningar lämnas ej in för dessa, om uträkningar ändå lämnas in kommer de inte att beaktas vid bedömningen.
- **Uppgift 6 – 7**: Svar med **FULLSTÄNDIGA REDOVISNINGAR** ska lämnas in.
  - Använd endast skrivpapper som tillhandahålls i skrivsalen.
  - För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
  - Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!
- Tentamen kan maximalt ge 60 + 40 = 100 poäng och för godkänt resultat krävs minst 50.
- Betygsgränser:

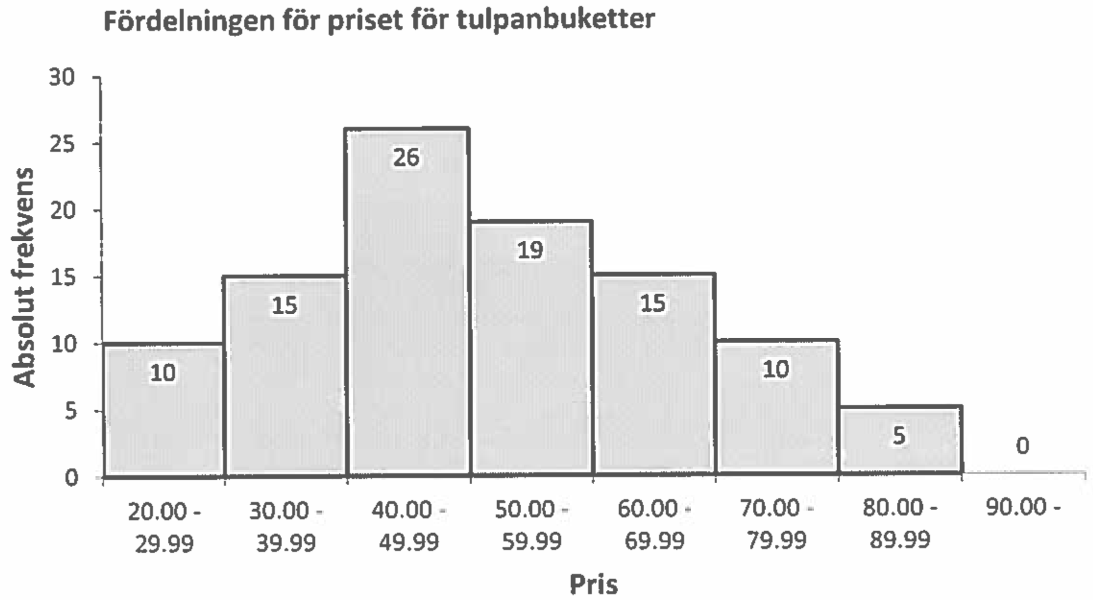
  A: 90 – 100 p
  B: 80 – 89 p
  C: 70 – 79 p
  D: 60 – 69 p
  E: 50 – 59 p
  Fx: 40 – 49 p
  F: 0 – 40 p

  OBS! Fx och F är underkända betyg som kräver omexamination. Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.
- Lösningsförslag läggs ut på Mondo kort efter tentamen.

LYCKA TILL!

Uppgift 1

En blomstergrossist analyserade dagspriset för färdigpackade tulpanbuketter som såldes till dagligvaruhandeln för de senaste 100 dagarna. Fördelningen över priset återges i diagrammet nedan.



a) Vilket av följande påståenden är <u>falskt</u>? (4p)

A. Medianen ligger i intervallet 40,00 – 49,99

B. Det är uppskattningsvis mer sannolikt att dagspriset för en slumpvist vald dag är mindre än 80 kr än att dagspriset är lika med eller större än 40 kr

C. Tredje kvartilen ligger i intervallet 60,00 – 69,99

D. Variationsvidden (*range*) är maximalt lika med 69,99

E. Fördelningen är vänstersned (*skewed-left*)

Samma blomstergrossit sammanställde statistik över pris och antal sålda tulpanbuketter till en viss kund för $n = 5$ slumpvist valda dagar. Den s.k. *variationskoefficienten* är ett mått på relativ spridning som mäter förhållandet i % mellan standardavvikelse och medelvärdet (förutsatt att medelvärdet är ett positivt tal) och beräknas som $cv(x) = 100 \cdot s_x/\bar{x}$.

Pris (x_i)	85	63	55	40	25
Antal sålda (y)	20	22	48	38	55

$\sum_i x_i = 268$ $\quad \sum_i (x_i - \bar{x})^2 = 2079{,}2$

$\sum_i y_i = 183$ $\quad \sum_i (y_i - \bar{y})^2 = 959{,}2$

$\sum_i x_i y_i = 8621$

b) Beräkna och ange $cv(x)$ för pris samt korrelationen r_{xy} mellan pris x och antal sålda y. (6p)

A. $cv(x) = 32{,}1\%$ $\quad r_{xy} = -0{,}841$

B. $cv(x) = 42{,}5\%$ $\quad r_{xy} = 0{,}841$

C. $cv(x) = 42{,}5\%$ $\quad r_{xy} = -0{,}841$

D. $cv(x) = 32{,}1\%$ $\quad r_{xy} = -0{,}673$

E. $cv(x) = 42{,}5\%$ $\quad r_{xy} = -0{,}673$

Uppgift 2

Ur en grupp på 200 patienter valdes 20 ut som fick en ny sorts medicinering (M) som enligt uppgift ska göra patienter friska (F) i 75 % av fallen jämfört med endast 50 % om de får den vanliga medicineringen ($\bar{M}$). Översatt till sannolikheter skulle vi skriva detta som

$$P(M) = 0{,}10 \qquad P(\bar{M}) = 0{,}90 \qquad P(F|M) = 0{,}75 \qquad P(F|\bar{M}) = 0{,}50$$

a) Vad är sannolikheten att en slumpvist vald patient blev frisk och fick den nya medicineringen? (4p)

A. 0,675
B. 0,075
C. 0,025
D. 0,375
E. 0,050

b) Vad är sannolikheten att en slumpvist vald patient blev frisk? (5p)

A. 0,425
B. 0,675
C. 0,625
D. 0,725
E. 0,525

Låt X vara en slumpvariabel med utfallsrummet $S_X = \{1,2,3,4,5\}$ och sannolikhetsfunktion

$$P(x) = \frac{2 \cdot (3 - x)^2 + 1}{25}$$

c) Använd att väntevärdet för X är 3 för att beräkna och ange

d) variansen σ_X^2 för X. (6p)

A. 3,12
B. 2,50
C. 2,00
D. 1,68
E. 3,28

Uppgift 3

Anta att slumpvariabeln X är normalfördelad med parametrarna $\mu_X = 20$ och $\sigma_X^2 = 25$ och att slumpvariabeln Y är normalfördelad med parametrarna $\mu_Y = 20$ och $\sigma_Y^2 = 100$. Anta att X och Y är oberoende.

a) Vilket av nedanstående alternativ är ett <u>sant</u> påstående? (6p)

A. $P(X > 20) \neq P(Y < 20)$

B. $P(X - Y > 0) \neq 0{,}5$ TIPS: Detta är en linjärkombination

C. $P(X > 5) = P(Y > 10)$

D. $P(X - 20 > 5) = P(Y - 20 > 10)$

E. Inget av alternativen A-D är korrekt

På ett fält i Sandemars naturreservat växer massor av kungsängsliljor. Cirka 80 % av blommorna är lila och 20 % är vita. Du plockar på måfå $n = 16$ stycken blommor och låter X beteckna antalet lila blommor.

b) Beräkna och ange sannolikheten $P(X \geq 12)$. (5p)

A. 0,798

B. 0,927

C. 0,598

D. 0,202

E. 0,918

OBS! Svarsalternativen har avrundats till 3 decimaler.

Betrakta ett stickprov av oberoende och lika fördelade (*iid*) slumpvariabler X_i för $i = 1, \dots, n$ där $X_i = 0$ eller 1 och sannolikheterna för utfallen är $P(X_i = 1) = P$ respektive $P(X_i = 0) = 1 - P$ där $0 < P < 1$. Låt sedan $W = \sum_{i=1}^{n} X_i$ och låt $\hat{p} = W/n$.

a) Vilket av följande alternativ är ett <u>falskt</u> påstående? (4p).

A. W är binomialfördelad enligt $Bin(n, P)$.

B. $\hat{p}$ är approximativt normalfördelad om stickprovet är tillräckligt stort

C. Standardavvikelsen för $\hat{p}$ är lika med $P(1-P)/n$.

D. $\hat{p}$ är en väntevärdesriktig (*unbiased*) skattning av P.

E. Centrala gränsvärdessatsen kan tillämpas i detta fall vad gäller sannolikhetsfördelningen för W om stickprovsstorleken är tillräckligt stor.

Uppgift 4

Ett företag tillverkar en känslig komponent som används i kärnkraftverk och som kräver särskild förpackning i samband med transport. I syfte att dra ner på fraktkostnaderna vill man testa en ny förpackningsmetod som ska dra ner den totala vikten. Man testar metoden genom att först packa enligt den gamla metoden och väger, sedan packar man om enligt den nya metoden och väger på nytt. Följande resultat för $n = 9$ komponenter erhölls där vikten anges i 1000 kg (ton):

	Komponent i	1	2	3	4	5	6	7	8	9	Σ
Vikt i 1000 kg	Äldre metod (x_i)	3,5	3,3	2,9	3,7	3,0	2,8	2,4	3,8	3,3	28,7
	Ny metod (y_i)	3,1	2,9	3,0	3,4	3,0	2,6	2,2	3,7	3,0	26,9
	Skillnad	0,4	0,4	-0,1	0,3	0,0	0,2	0,2	0,1	0,3	1,8
	Skillnad i kvadrat	0,16	0,16	0,01	0,09	0,00	0,04	0,04	0,01	0,09	0,6

Eftersom den nya metoden är lite dyrare och den lönar sig bara om den genomsnittliga viktminskningen är mer än 100 kg (dvs. 0,1 ton), så vill du testa på $\alpha = 5$ % signifikansnivå om

H_0: genomsnittlig skillnad = 0,1 mot H_1: genomsnittlig skillnad > 0,1

Du antar att vikterna är *iid* för respektive förpackningsmetod och normalfördelade.

a) Ange det kritiska värdet för testet. (4p)

A. kritiskt värde = 1,645
B. kritiskt värde = 1,960
C. kritiskt värde = 1,860
D. kritiskt värde = 1,833
E. kritiskt värde = 2,306

b) Ange det observerade värdet på testvariabeln och den korrekta slutsatsen. TIPS: Enklast är att utgå från summorna i tabellen ovan när du gör dina beräkningar. (6p)

A. observerat värde = 1,732 och H_0 förkastas inte
B. observerat värde = 1,732 och H_0 förkastas
C. observerat värde = 0,481 och H_0 förkastas inte
D. observerat värde = 3,464 och H_0 förkastas
E. observerat värde = 3,464 och H_0 förkastas inte

Uppgift 5

I samband med en utredning om social trygghet genomfördes en urvalsundersökning där man bland annat ställde frågor om man upplever att man utsätts för diskriminering i vardagen pga. av kön, ålder eller utländsk bakgrund. Svarsalternativen var på ordinalskala med fem alternativ men då antalet svarande var för få slog man ihop de fem klasserna till två: " ofta eller ibland" och "sällan eller aldrig". Följande tabell redovisar svaren efter vilken bakgrund de svarande hade:

Frekvens	Ofta el. Ibland	Sällan el. aldrig	Totalt
Svenska föräldrar	24	36	60
Minst en utlandsfödd förälder	21	9	30
Totalt	45	45	90

Du vill genomföra ett test för att se om bakgrund och hur ofta man upplever diskriminering i vardagen är oberoende eller beroende. Utgå ifrån signifikansnivån $\alpha = 1$ %.

a) Ange den korrekta slutsatsen av testet. (7p)

A. $\chi^2_{\text{obs}} = 8{,}21 > \chi^2_{\text{krit}} = 6{,}635$ och H_0 förkastas på 1 % nivån

B. $\chi^2_{\text{obs}} = 7{,}20 < \chi^2_{\text{krit}} = 7{,}879$ och H_0 förkastas inte på 1 % nivån

C. $\chi^2_{\text{obs}} = 7{,}20 > \chi^2_{\text{krit}} = 6{,}635$ och H_0 förkastas inte på 1 % nivån

D. $\chi^2_{\text{obs}} = 7{,}20 > \chi^2_{\text{krit}} = 6{,}635$ och H_0 förkastas på 1 % nivån

E. $\chi^2_{\text{obs}} = 8{,}21 > \chi^2_{\text{krit}} = 7{,}879$ och H_0 förkastas på 1 % nivån

En tidsserie brukar ibland beskrivas som en summa, eller produkt, av olika *komponenter* som var för sig avspeglar olika egenskaper i tidsserien man studerar. Olika metoder används för att i olika syften bearbeta en tidsserie för analys och vid behov för prediktioner.

b) Ange vilket av följande påstående som <u>inte</u> är korrekt. (3p)

A. Man kan skatta trenden i en serie med glidande medelvärdesmetoden.

B. Man kan skatta trenden i en serie med regressionsanalys med tiden som beroende variabel.

C. Säsongsvariationen kan modelleras med dummyvariabler i en regressionsmodell.

D. Med en multiplikativ modell säsongsrensar man genom att dela det observerade värdet med den skattade säsongsfaktorn.

E. Med en additiv modell säsongsrensar man genom att subtrahera säsongsfaktorn från det observerade värdet.

Fullständig redovisning krävs för följande uppgifter.

Använd separata pappersark för uppgift 6 resp. uppgift 7.

För uppgift 7c används Svarsbilagan

Uppgift 6

För universitetslärare och forskare registreras s.k. publiceringspoäng när de får sin forskning publicerad, detta som ett mått på forskningsaktiviteten både individuellt och för institutionen som helhet. I tabellen nedan presenteras deskriptiv statistik över $X =$ publiceringspoängen från det senaste året och $Y =$ den relativa löneförändringen i % vid senaste lönerevision för $n = 12$ universitetslärare. De 12 observationerna har plottats i ett spridningsdiagram som finns bifogad till svarsbilagan.

Variabel	N	Mean	StDev	Min	Q1	Median	Q3	Max
Publicerings-poäng (X)	12	2,500	3,656	0,00	0,00	0,50	4,00	12,00
Procentuell löne-förändring (Y)	12	2,381	0,7904	1,363	2,035	2,256	2,420	4,688

a) Illustrera respektive variabel med varsin boxplot. Obs! För variabel Y är statistiken angiven med tre decimalers noggrannhet men du behöver inte vara så exakt när du ritar boxploten. (6 poäng)

b) Beräkna ett 95 % konfidensintervall för $\mu_Y =$ den genomsnittliga löneförändringen i populationen. Ange tydligt förutsättningar och antaganden som krävs, formel, formel med insatta värden och slutresultat. Tolka sedan resultatet. (10 poäng).

c) Med ledning endast av konfidensintervallet i b)-uppgiften, skulle du kunna förkasta $H_0: \mu_Y = 2{,}5$ till förmån för $H_1: \mu_Y \neq 2{,}5$? Förklara kortfattat! Även om du inte gjort uppgift b) kan du få poäng för ett korrekt resonemang för hur du skulle kunna avgöra frågan. (4p)

Uppgift 7

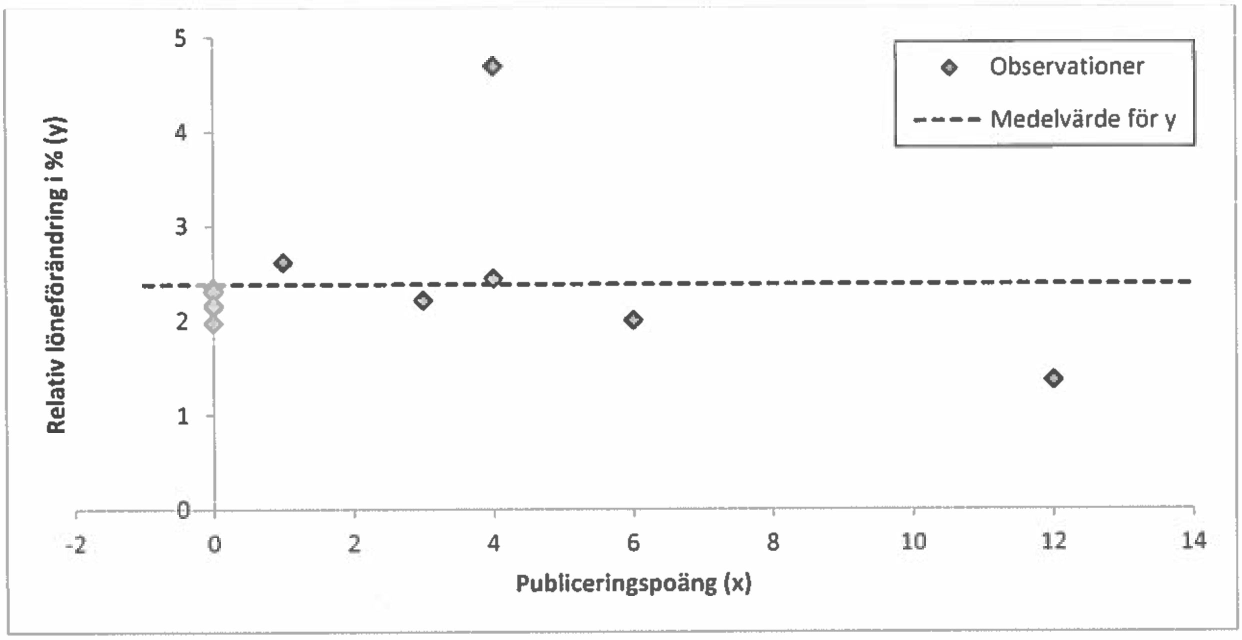
Denna uppgift är en fortsättning från uppgift 6 ovan. För att studera hur den relativa löneförändringen (Y) påverkas av publiceringspoängen (X) har variablerna plottats mot varandra, se digrammet på nästa sida. Samma diagram har även bifogats svarsbilagan.

a) Skatta parametrarna β_0 och β_1 i modellen $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ med minsta kvadratmetoden och tolka sedan parameterskattningarna. OBS! För att skatta modellen behöver du bara de deskriptiva resultaten som gavs i uppgift 6 ovan samt summan $\sum_i x_i y_i = 66{,}139$. (10p)

b) Säg att du vill testa $H_0: \beta_1 = 0$ mot $H_1: \beta_1 > 0$ på signifikansnivån 5 %. Om den skattade lutningen är negativ, vad blir din slutsats? Basera din slutsats på ett kortfattat resonemang kring vilka möjliga p-värden du skulle kunna få när den skattade lutningen är negativ. (5p).

c) I spridningsdiagrammet finns en observation som kommer att ge upphov till en särskilt stor residual vid regressionsanpassningen och en punkt som har stort inflytande på regressionslinjens lutning. Vilka punkter är det frågan om i respektive fall? Kort motivering krävs för full poäng. Markera respektive punkt i figuren som ligger ihop med svarsbilagan, skriv ditt svar under figuren och lämna sedan in den med övriga blad. (5p)

Spridningsdiagram för uppgift 7



Obs! Använd svarsbilagan för att besvara uppgift 7c

STOCKHOLMS UNIVERSITET
Statistiska institutionen

VT2017

SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer 2017-04-20

Skrivsal: ______________________________

Anonymkod: ______________________________ *(skriv tydligt!)*

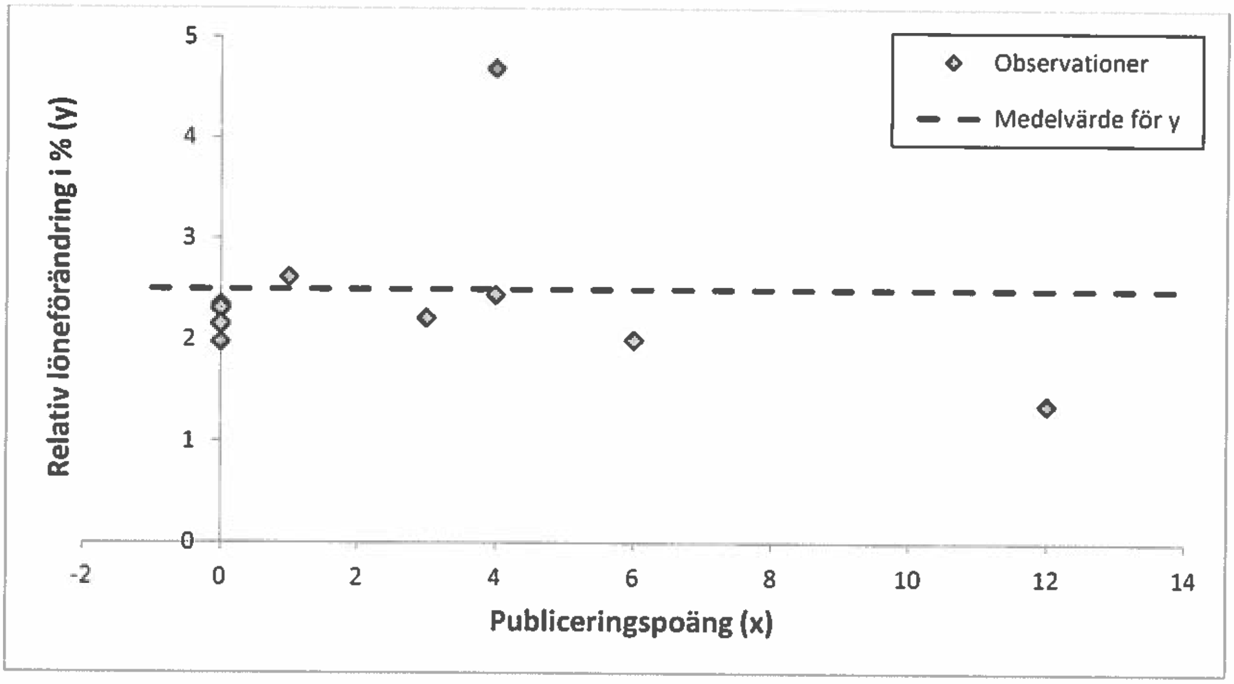
Markera ditt svar med ett tydligt kryss (X) i rutorna nedan.
OBS! Endast ett kryss per uppgift. Har fler än ett svarsalternativ markerats ges noll poäng.

OBS! Om du efter att ha kontrollerat dina beräkningar ordentligt kommer fram till att svaret inte finns bland de angivna svarsalternativen, skriv ditt svar i mariginalen till höger.

		A	B	C	D	E
Uppgift 1	a)	☐	☐	☐	☐	☐
	b)	☐	☐	☐	☐	☐
Uppgift 2	a)	☐	☐	☐	☐	☐
	b)	☐	☐	☐	☐	☐
	c)	☐	☐	☐	☐	☐
Uppgift 3	a)	☐	☐	☐	☐	☐
	b)	☐	☐	☐	☐	☐
	c)	☐	☐	☐	☐	☐
Uppgift 4	a)	☐	☐	☐	☐	☐
	b)	☐	☐	☐	☐	☐
Uppgift 5	a)	☐	☐	☐	☐	☐
	b)	☐	☐	☐	☐	☐

Svarsbilaga till uppgift 7c

Spridningsdiagram



Skriftliga kommentarer och svar till uppgift 7c kan ges här:

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)

VT2017

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2017-04-20

LÖSNINGSFÖRSLAG

Första version, med reservation för tryck- och slarvfel / 2017-04-20 MC

Sammanfattning SVARSBILAGA Uppgifter 1-5
Utförliga beräkningar ges på efterföljande sidor

		A	**B**	**C**	**D**	**E**
Uppgift 1	a)	□	□	□	□	■
	b)	□	□	■	□	□
Uppgift 2	a)	□	■	□	□	□
	b)	□	□	□	□	■
	c)	■	□	□	□	□
Uppgift 3	a)	□	□	□	■	□
	b)	■	□	□	□	□
	c)	□	□	■	□	□
Uppgift 4	a)	□	□	■	□	□
	b)	■	□	□	□	□
Uppgift 5	a)	□	□	□	■	□
	b)	□	■	□	□	□

Uppgift 1

a) Rätt svar: **E**

A. Beräkna $(100+1)\cdot 0{,}50 = 50{,}5$; medianen definieras alltså som mittpunkten mellan 50:e och 51:a observationen; summera frekvenserna från vänster så inses att båda dessa observationer ligger i intervallet 40,00 – 49,99. Sant påstående.

B. $P(pris < 80) =$ summera frekvenserna upp till och med 70,00 – 79,99 så har man 95 av 100; $P(pris > 40) =$ summera frekvenserna från 40,00 – 49,99 så har man 75 av hundra. Sant påstående.

C. Beräkna $(100+1)\cdot 0{,}75 = 75{,}75$; resonera på samma sätt som i A så inses att tredje kvartilen ligger i intervallet 60,00 – 69,99. Sant påstående.

D. Variationsvidden (*range*) definieras som $Max - Min$; maxvärdet kan inte vara större än 89,99 och minimivärdet kan inte vara mindre än 20; max *range* är alltså 89,99 – 20 = 69,99. Sant påstående.

E. Fördelningen är inte vänstersned (*skewed-left*) eftersom högersvansen är mer utdragen jämfört med vänstersidan; den är högersned; se t.ex. NCT sid 45. **FALSKT påstående.**

Se kurslitteraturen och föreläsningsanteckningar.

b) Rätt svar: **C**

$$\bar{x} = \frac{268}{5} = 53{,}6 \qquad s_x^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{2079{,}2}{4} = 519{,}8 \qquad s_x = 22{,}799$$

(1) $$cv(x) = 100\cdot\frac{s_x}{\bar{x}} = 100\cdot\frac{22{,}799}{53{,}6} = \mathbf{42{,}5\%}$$

$$\bar{y} = \frac{183}{5} = 36{,}6 \qquad s_y^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1} = \frac{959{,}2}{4} = 239{,}8 \qquad s_y = 15{,}485$$

$$s_{xy} = Cov(x,y) = \frac{\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}}{n-1} = \frac{8621 - 5\cdot 53{,}6\cdot 36{,}6}{4} = \frac{-1187{,}8}{4} = -296{,}95$$

(2) $$r_{xy} = Corr(x,y) = \frac{s_{xy}}{s_x\cdot s_y} = \frac{-296{,}95}{22{,}799\cdot 15{,}485} = \mathbf{-0{,}841}$$

Uppgift 2

Givet: $P(M) = 0{,}10$ $\quad P(\bar{M}) = 0{,}90$ $\quad P(F|M) = 0{,}75$ $\quad P(F|\bar{M}) = 0{,}50$

a) Rätt svar: **B**

Sökt: $P(M\cap F) =$ [multiplikationssatsen] $= P(F|M)\cdot P(M) = 0{,}75\cdot 0{,}10 = \mathbf{0{,}075}$

b) Rätt svar: **E**

Sökt: $P(F) =$ [satsen om total sannolikhet] $= P(F|M)\cdot P(M) + P(F|\bar{M})\cdot P(\bar{M})$

$$= 0{,}75\cdot 0{,}10 + 0{,}50\cdot 0{,}90 = \mathbf{0{,}525}$$

c) Rätt svar: **A**

Variansen ges av $\sigma_X^2 = \sum_{x \in S_x} (x - \mu_X)^2 P(x) = \sum_{x \in S_x} x^2 P(x) - \mu_X^2$.

Beräkna först sannolikheterna för alla x i utfallsrummet $\{1,2,3,4,5\}$, sedan beräknas $(x - \mu_X)^2 = (x - 3)^2$ (eller x^2 om man föredrar det) för alla x, multiplicera parvis och summera:

x	1	2	3	4	5	Summa
$P(x)$	$9/25$	$3/25$	$1/25$	$3/25$	$9/25$	1
$(x-3)^2$	4	1	0	1	4	
$P(x) \cdot (x-3)^2$	$36/25$	$3/25$	0	$3/25$	$36/25$	$78/25 = \mathbf{3,12}$

Uppgift 3

a) Rätt svar: **D**

A. $P(X > 20) = P\left(Z > \frac{20-20}{5}\right) = P(Z > 0) = 0{,}5;$

$P(Y < 20) = P\left(Z < \frac{20-20}{10}\right) = P(Z < 0) = 0{,}5$; de är lika; <u>inte</u> ett korrekt påstående.

B. $P(X - Y > 0) =$ [linjärkombination] $= P\left(Z > \frac{(20-20)-0}{\sqrt{25+100}}\right) = P(Z > 0) = 0{,}5;$

sannolikheten är lika med 0,5; <u>inte</u> ett korrekt påstående.

C. $P(X > 5) = P\left(Z > \frac{5-20}{5}\right) = P(Z > -3)$

$P(Y > 10) = P\left(Z > \frac{10-20}{10}\right) = P(Z > -1)$; de är inte lika; <u>inte</u> ett korrekt påstående.

D. $P(X - 20 > 5) = P(X > 25) = P\left(Z > \frac{25-20}{5}\right) = P(Z > 1)$

$P(Y - 20 > 10) = P(Y > 30) = P(Z > \frac{30-20}{10} = P(Z > 1);$

de är lika; **SANT påstående.**

E. Eftersom D är ett korrekt påstående gäller att E <u>inte</u> är ett korrekt påstående.

b) Rätt svar: **A**

Givet att $X \sim Bin(16; 0{,}8)$. Eftersom tabellen inte innehåller värden för $P > 0{,}5$ definierar vi istället $16 - X = Y \sim Bin(16; 0{,}2)$.

Sökt: $P(X \geq 12) = P(n - X \leq n - 12) = P(Y \leq 4) =$ [enligt Tabell 7] $= \mathbf{0,79825}$

Man kan ”se” lösningen genom att lista utfallen för X och motsvarande utfall för Y och ringa in de gynnsamma utfallen:

X:	...	10	11	12	13	14	15	16
						$\updownarrow$		
$16 - X = Y$:	...	6	5	4	3	2	1	0

c) Rätt svar: **C**

A. W är summan av n oberoende 0-1 variabler var och en med sannolikhet P att få utfallet 1; detta är definitionen av binomialfördelad slumpvariabel. Sant påstående.

B. $\hat{p}$ är definierad som medelvärdet av n stycken *iid* slumpvariabler; enligt CGS är denna då approximativt normalfördelad om stickprovet är tillräckligt stort. Sant påstående.

C. Variansen för $\hat{p}$ kan beräknas enligt

$$Var(Y/n) = (1/n^2) \cdot Var(Y) = (1/n^2) \cdot nP(1-P) = P(1-P)/n$$

Standardavvikelsen är kvadratroten ur detta. **FALSKT påstående**.

D. Väntevärdet kan beräknas $E(\hat{p}) = E(Y/n) = (1/n)E(Y) = (1/n)nP = P$, alltså en väntevärdesriktig skattning av P. Sant påstående.

E. Eftersom $W = n\hat{p}$ och $\hat{p}$ är approximativt normalfördelad enligt CGS så gäller detsamma för W. Sant påstående.

Se kurslitteraturen och föreläsningsanteckningar.

Uppgift 4

Notera att det är samma $n = 9$ komponenter som man prövar de två förpackningsmetoderna på, alltså är det ett *parvist beroende*.

Låt $D = X - Y$. Eftersom X och Y båda är (*iid*) normalfördelade är D garanterat normalfördelad. Beteckna den genomsnittliga differensen i populationen med μ_D, skattas med $\bar{d} = 1{,}8/9 = 0{,}2$. Variansen betecknas med σ_D^2 och är okänd och skattas med

$$s_d^2 = \frac{\sum_{i=1}^{n} d_i^2 - n\bar{d}^2}{n-1} = \frac{0{,}6 - 9 \cdot 0{,}2^2}{8} = 0{,}03$$

Hypoteserna är H_0: $\mu_D = 0{,}1$ mot H_1: $\mu_D > 0{,}1$ och under dessa förutsättningar är testvariabeln

$$t = \frac{\bar{d} - 0{,}1}{\sqrt{0{,}03/9}} \sim t\text{-fördelad med } n-1 = 8 \text{ frihetsgrader}$$

a) Rätt svar: **C**

Kritisk gräns för det enkelsidiga testet är $t_{krit} = t_{8;0{,}05} = \mathbf{1{,}860}$ om $\alpha = 0{,}05$ används

b) Rätt svar: **A**

Enkelsidigt test till höger medför att man förkastar H_0 om $t_{obs} > t_{krit}$. Insättning ger

$$t_{obs} = \frac{0{,}2 - 0{,}1}{\sqrt{0{,}03/9}} = 1{,}732 < 1{,}860 \;\Rightarrow\; H_0 \text{ kan inte förkastas}$$

Uppgift 5

a) Rätt svar: **D**

Oberoende test, ett χ^2-test. Man utgår ifrån hypoteserna

H_0: variablerna är oberoende mot H_1: variablerna är beroende

Man har $r = c = 2$ rader och kolumner vilket ger att testvariabeln är χ^2-fördelad med $(r-1)(c-1) = 1$ frihetgrad. Testvariabeln är

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \quad \text{där}$$

där O_{ij} är de observerade frekvenserna och $E_{ij} = R_i C_j / n$ är de förväntade frekvenserna under H_0. Beslutsregel är att man förkastar H_0 om

$$\chi^2_{\text{obs}} > \chi^2_{\text{krit}} = \chi^2_{1;0,01} = \text{[enligt Tabell 4]} = \mathbf{6,635}$$

Beräkningar ger:

O_{ij} = observerat

	Ofta	Sällan	Totalt
Sv.	24	36	60
Utl.	21	9	30
Totalt	45	45	90

E_{ij} = förväntat

	Ofta	Sällan	Totalt
Sv.	30	30	60
Utl.	15	15	30
Totalt	45	45	90

$(O_{ij}-E_{ij})^2/E_{ij}$

	Ofta	Sällan
Sv.	1,2	1,2
Utl.	2,4	2,4

$\chi^2_{\text{obs}} = \Sigma\Sigma =$ **7,2**

Slutsats: $\chi^2_{\text{obs}} = 7{,}20 > \chi^2_{\text{krit}} = 6{,}635$ och H_0 förkastas på 1 % signifikansnivå.

b) Rätt svar: **B**

A. Man kan skatta trenden i en serie med glidande medelvärdesmetoden. Sant.

B. Man kan skatta trenden i en serie med regressionsanalys med tiden som beroende variabel. **FALSKT**, man använder tiden som oberoende variabel, dvs. förklaringsvariabel

C. Säsongsvariationen kan modelleras med dummyvariabler i en regressionsmodell. Sant.

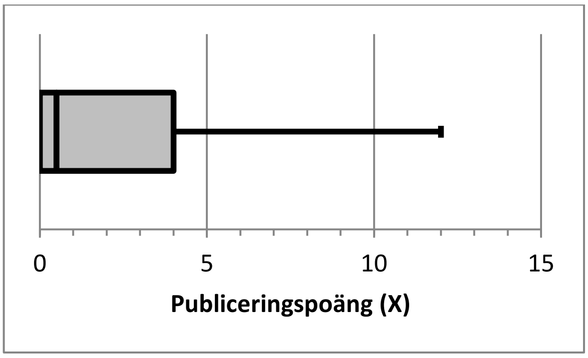
D. Med en multiplikativ modell säsongsrensar man genom att dela det observerade värdet med den skattade säsongsfaktorn. Sant.

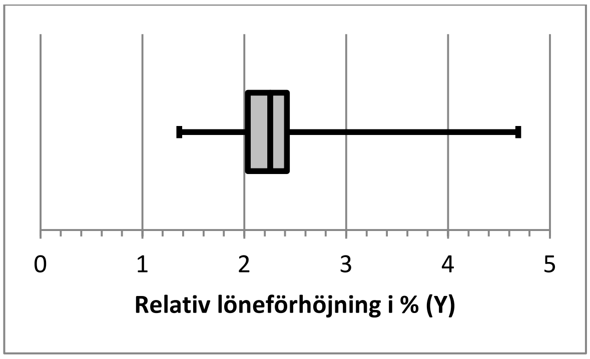
E. Med en additiv modell säsongsrensar man genom att subtrahera säsongsfaktorn från det observerade värdet. Sant.

Se kurslitteraturen och föreläsningsanteckningar.

Uppgift 6

a) En boxplot visar grafiskt läget för fem-siffer-sammanfattningen (five-number-summary): minimum, första kvartilen, medianen, tredje kvartilen. Samtliga dessa värden är givna. Eftersom det är olika skalor kan det vara lämpligt med två olika diagram. Kan göras stående eller liggande. Notera att för X-variabeln sammanfaller min och Q_1 pga. att antalet med noll publiceringspoäng är stort (faktiskt sex stycken, se spridningsdiagrammet i uppgift 7).





b) <u>Antaganden</u>: alla Y_i är normalfördelade med samma väntevärde μ_Y och varians σ_Y^2 och samtliga observationer är oberoende av varandra (iid). Vidare är σ_Y^2 okänd så den måste skattas med s_y^2 och vi måste använda t-fördelningen.

<u>Formel</u>: ett 95 % konfidensintervall ($\alpha = 0{,}05$) för μ_Y ges av: $\bar{y} \pm t_{n-1;\, \alpha/2} \cdot \frac{s_y}{\sqrt{n}}$

där $t_{n-1;\, \alpha/2} = t_{11;\, 0{,}025} =$ [enligt Tabell 3] $= 2{,}201$

<u>Insättning</u> ger

$$2{,}381 \pm 2{,}201 \cdot \frac{0{,}7904}{\sqrt{12}} = \mathbf{2{,}381 \pm 0{,}502} \qquad \text{eller} \qquad (\mathbf{1{,}879}\,;\ \mathbf{2{,}883})$$

<u>Slutsats/tolkning</u>: Med 95 % konfidens (ej sannolikhet) ligger den sanna genomsnittliga relativa löneökningen μ_Y i konfidensintervallet. Tolkningen av konfidensgraden är att vid upprepade stickprovsdragningar kommer det sanna värdet μ_Y ligga i 95 % av konfidensintervall man kan få, allt annat lika.

c) Om man vill genomföra ett dubbelsidigt hypotestest, dvs. $H_0\!: \mu_Y = \mu_0$ mot $H_1\!: \mu_Y \neq \mu_0$ för något tal μ_0 på signifikansnivå α, så räcker det att se om detta tal ligger i motsvarande $100(1-\alpha)$ % konfidensintervall. Konfidensintervallet kan sägas innehålla alla de värden som skulle medföra att vi inte kan förkasta nollhypotesen. I detta fall är $\mu_0 = 2{,}5$ och detta värde ligger i intervallet och vi skulle alltså inte förkasta H_0.

Uppgift 7

a) *Ej klart, fullständig lösning kommer inom kort.*

Skattningarna blir:

Intercept (konstant): $b_0 = 2{,}471$

Lutningkoefficient: $b_1 = -0{,}03599$

b) *Ej klart, fullständig lösning kommer inom kort, men lite kortfattat:*

Eftersom testet är enkelsidigt till höger så är den kritiska gränsen positiv, $t_{\text{krit}} = t_{n-2;\,\alpha}$ och man förkastar H_0 om man ligger till höger om den punkten, dvs. om $t_{\text{obs}} > t_{\text{krit}}$ och man förkastar inte om man ligger till vänster, dvs. om $t_{\text{obs}} < t_{\text{krit}}$.

För detta enkelsidiga test beräknas p-värdet som den del av arean under ”t-klockan” som ligger till höger om det observerade värdet på testvariabeln, $t_{\text{obs}} = b_j/s_{b_j}$.

Men om skattningen b_1 är negativ så måste även testvariabeln t_{obs}, bli negativ. Detta medför dels att man alltid kommer att förkasta detta test och dels att arean till höger om t_{obs} och under klockan, dvs. p-värdet minst måste vara 0,5.

Bild kommer

c) *Ej klart, fullständig lösning kommer inom kort.*

Punkterna/observationerna är:

Den som ligger högst upp i mitten av spridningsdiagrammer, den ger en stor residual.

Den som ligger längst till höger, ger stort inflytande på parameterskattningarna.