

# Løsningsforslag eksamen – Høgskolen i Østfold

SFB10711 – Statistikk

Våren 2019

## 1

Boksalget ( $\mathbf{X}$ , oppgitt i antall tusen bøker solgt) til en av deres mest populære forfattere (Wiktor De Brune) de siste årene vises i tabellen nedenfor:

	2014	2015	2016	2017	2018
$\square$	150	140	160	120	180

a) Hva er gjennomsnittet ( $\bar{X}$ ), variansen ( $S_X^2$ ) og standardavviket ( $S_X$ ) for disse inntektene?

b) Beregn et 95% konfidensintervall for  $\bar{X}$ .

---

$$\bar{X} = \frac{150 + 140 + 160 + 120 + 180}{5} = 150 \quad (1)$$

$$S_X^2 = \frac{(150 - 150)^2 + (140 - 150)^2 + (160 - 150)^2 + (120 - 150)^2 + (180 - 150)^2}{4} = 500 \quad (2)$$

$$S_X = \sqrt{500} = 22,36 \quad (3)$$

$$t_{0,05/2}^{5-1} = 2,776 \quad (4)$$

$$S[\bar{X}] = \frac{S_X}{\sqrt{5}} = 10 \quad (5)$$

$$\bar{X} = 150 \pm 2,776 * 10 = [122,24 \bullet 177,76] \quad (6)$$

---

Dere har så sett på salget til en annen kjent forfatter i samme sjanger (Heidi Kurås) i samme perioder ( $\mathbf{Y}$ ). Hennes tall er 100, 130, 120, 150 og 100.  $\bar{Y} = 120$  og  $S_Y^2 = 450$ .

c) Beregn korrelasjonskoeffisienten  $R_{XY}$  for boksalget til De Brune ( $\mathbf{X}$ ) og Kurås ( $\mathbf{Y}$ ).

d) Hva kan du si om forholdet mellom salget til de to forfatterne, når ledelsen spør deg om hva disse tallene egentlig betyr?

---

$$S_{XY} = \frac{1}{5-1}((150 - 150)(100 - 120) + (140 - 150)(130 - 120) + (160 - 150)(120 - 120) \quad (7)$$

$$+ (120 - 150)(150 - 120) + (180 - 150)(100 - 120)) = \frac{-1600}{4} = -400 \quad (8)$$

$$R_{XY} = \frac{-400}{22,36 * 21,21} = -0,843 \quad (9)$$

---

## 2

Du jobber i forlaget God Bok AS. Dere har fulgt salget på noen utvalgte bøker over tid, og for hver måned er det 30% sjans for at boken selger 150 eksemplarer, 40% sjans for at dere selger 250 eksemplarer, og 30% sjans for at dere selger 350 eksemplarer.

a) Hva er forventet antall bøker solgt per måned ( $E[\mathbf{X}]$ ), og hva er variansen og standardavviket til  $\mathbf{X}$ ?

$$E[X] = 150 * 0,3 + 250 * 0,4 + 350 * 0,3 = 250 \quad (10)$$

$$Var[X] = (150 - 250)^2 * 0,3 + (250 - 250)^2 * 0,4 + (350 - 250)^2 * 0,3 = 6000 \quad (11)$$

$$E[X^2] = (150^2) * 0,3 + (250^2) * 0,4 + (350^2) * 0,3 = 68500 \quad (12)$$

$$Var[X] = 68500 - 250^2 = 6000 \quad (13)$$

$$\sigma[X] = \sqrt{6000} = 77,46 \quad (14)$$

Ledelsen ønsker å eksperimentere med prisingen av bøker som er eldre enn et år, og har over tid gjennomført et eksperiment på et sett med sammenlignbare bøker. Prisen varierte mellom 150 kroner, 175 kroner eller 200 kroner. La prisen være  $Y$ . Mengden bøker som ble solgt ( $X$ ) i hver periode var 150 bøker, 250 bøker eller 350 bøker.

### Simultansannsynligheter

		Mengde (X)			
Pris (Y)	150	250	350		
150	3,33 %	13,33 %	23,33 %		40 %
175	10 %	16,67 %	3,33 %		30 %
200	16,67 %	10 %	3,33 %		30 %
	30 %	40 %	30 %		100.0%

Figur 1: Simultan sannsynlighetsfordeling

Oppgitte verdier:  $E[Y] = 172,5$  og  $Var[Y] = 431,25$ .

Følgende tabell gir funksjonsverdiene for  $X \cdot Y$ :

		Funksjonsverdier X*Y		
	150	250	350	
150	22 500	37 500	52 500	
175	26 250	43 750	61 250	
200	30 000	50 000	70 000	

Figur 2: Simultan sannsynlighetsfordeling

b) Regn ut korrelasjonskoeffisienten  $\rho[X, Y]$ .

$$E[XY] = 22500 * 0,033 + 37500 * 0,133 + 52500 * 0,233 + \quad (15)$$

$$26500 * 0,1 + 43750 * 0,167 + 61250 * 0,033 + \quad (16)$$

$$30000 * 0,167 + 50000 * 0,1 + 70000 * 0,033 = 42260 \quad (17)$$

$$Cov[X, Y] = 42260 - 250 * 172,5 = -865 \quad (18)$$

$$\rho[X, Y] = \frac{-865}{77,46 * 173,74} = -0,53 \quad (19)$$

## 3

Dere får hvert år et stort antall manus tilsendt, og 15% av disse er gode nok til å bli sendt til vurdering. I løpet av påskeuken fikk dere tilsendt 20 manus som dere nå skal gå gjennom.

a) Hva er sannsynligheten for at ingen av disse er gode nok til å bli sent til vurdering ( $P(X = 0)$ )?

b) Hva forventet antall bøker som blir sendt til vurdering ( $E[X]$ ), og hva er variansen til  $X$ ?

$$P(X = 0) = \binom{20}{0} 0,15^0 * 0,85^{20} = 0,039 \quad (20)$$

$$E[X] = 20 * 0,15 = 3 \quad (21)$$

$$Var[X] = 20 * 0,15 * 0,85 = 2,55 \quad (22)$$

---

Av alle manusene dere får inn, er det kun 2% som ender med utgivelse. Dere forventer å få inn ca. 150 manus i år. (Hint: liten p og stor n.)

c) Hva er forventet antall manus som ender med utgivelse i år ( $E[X]$ ), og hva er variansen til  $X$ ?

d) Hva er sannsynligheten for at maksimalt 2 av disse manuskriptene ender med utgivelse ( $P(X \leq 2)$ )?

---

$$\lambda = n * p = 150 * 0,02 = 3 \quad (23)$$

$$E[X] = \lambda = 3 \quad (24)$$

$$Var[X] = \lambda = 3 \quad (25)$$

$$P(X \leq 2) = P(0) + P(1) + P(2) \quad (26)$$

$$P(X = 0) = \frac{3^0}{0!} e^{-3} = 0,05 \quad (27)$$

$$P(X = 1) = \frac{3^1}{1!} e^{-3} = 0,149 \quad (28)$$

$$P(X = 2) = \frac{3^2}{2!} e^{-3} = 0,224 \quad (29)$$

$$P(X \leq 3) = 0,05 + 0,149 + 0,224 = 0,423 \quad (30)$$

$$(31)$$

---

## 4

40% av deres forfattere er kvinner. Dere utgir en del selvhjelpsbøker, og 60% av bøkene dere utgir i denne kategorien er skrevet av kvinner. Av alle bøkene dere utgir, er 30% selvhjelpsbøker.

a) Hva er sannsynligheten for at en av deres bøker er en selvhjelpsbok, dersom den er skrevet av en kvinne?

---

$$P(S|K) = P(K|S) * P(S) / P(K) \quad (32)$$

$$P(S|K) = 0,6 * 0,3 / 0,4 = 0,45 \quad (33)$$

---

Krimbøker er deres bestselgere, og 60% av bøkene dere utgir er krimbøker. Sannsynligheten for at en av deres krimbøker har en mannlig forfatter er 50%.

b) Hva er sannsynligheten for at en tilfeldig valgt bok er en krimbok som har en mannlig forfatter?

---

$$P(M|K) = \frac{P(M \cap K)}{P(K)} \quad (34)$$

$$0,5 = \frac{P(M \cap K)}{0,6} \quad (35)$$

$$0,5 * 0,6 = P(M \cap K) \quad (36)$$

$$P(M \cap K) = 0,3 \quad (37)$$

---

30% av selvhjelpsbøkene dere utgir selger mer enn 5000 eksemplarer. 80% av krimbøkene dere utgir selger mer enn 5000 eksemplarer. Av de øvrige bøkene (10% av bøkene deres) er sannsynligheten for å selge mer enn 5000 eksemplarer 15%. Se de tidligere oppgavene for informasjon om andelen bøker av forskjellige typer.

c) Hvor stor andel av bøkene dere utgir selger mer enn 5000 eksemplarer?

---

$$P(5000+) = 0,6 * 0,8 + 0,3 * 0,3 + 0,1 * 0,15 = 0,585 \quad (38)$$

---

## 5

Du har nå blitt sendt på bokmesse, og du har fått ansvar for å sette opp standen deres. Du har fått ansvar for å velge farge på standen, type dekorasjon, om innredning. Firmaet deres bruker tre farger, og dette er fargealternativene dine. Du har videre valget mellom å dekorere med blomster, bokstabler eller interaktive skjermer. Til sist blir du bedt om å velge mellom en innredning med høye bord og noen krakker, eller en litt mer uformell stue-løsning med sofa.

a) Hvor mange forskjellige varianter at standen deres kan du lage ut fra disse alternativene?

---

$$Kombinasjoner = 3 * 3 * 2 = 18 \quad (39)$$

På deres stand på bokmessen har dere 5 likeverdige utstillingsplasser som kan brukes til å profilere nye bøker. Du har med deg 15 bøker i kofferten, og må nå vurdere hvilke bøker som skal velges ut.

a) Hvor mange kombinasjoner av profilerte bøker finnes?

---

$$Kombinasjoner = \binom{15}{5} = 3003 \quad (40)$$

Et av målene dine på denne messen er å finne utenlandske suksesser dere kan utgi i Norge. Du har blitt bedt om å velge ut tre bøker, og prioritere disse slik at nummer 1 er boken dere forsøker å kjøpe først, deretter nummer 2, og så nummer 3. Det er i alt 15 bøker som er aktuelle for innkjøp.

b) Hvor mange forskjellige prioriteringslister kan du lage?

---

$$Permutasjoner = (15)_3 = 2730 \quad (41)$$

## 6

Dere skal gjennomføre en spørreundersøkelse, og som en del av forberedelsene har dere testet spørreskjemaet på 20 personer. Dere spurte bl.a. respondentene om hvor mange bøker de hadde lest i år. Gjennomsnittet  $\bar{X}$  ble 2,85 bøker, og det var 20 respondenter.  $S_X^2$  var 13,197.

a) Lag et 95% konfidensintervall for gjennomsnittet  $\bar{X}$ .

---

$$t_{0,05/2}^{(20-1)} = 2,093 \quad (42)$$

$$S[\bar{X}] = \frac{\sqrt{13,197}}{\sqrt{20}} \quad (43)$$

$$KI = 2,85 \pm (2,093 * 0,8123) = [1,15 \bullet 4,55] \quad (44)$$

Senere sendte dere ut spørreundersøkelsen til et større utvalg, og nå ble gjennomsnittet 3,5. Dere vet at standardavviket i populasjonen  $\sigma$  er 5. 150 personer ble nå undersøkt.

b) Lag et nytt 95% konfidensintervall for gjennomsnittet  $\bar{X}$ .

---

$$z_{0,05/2} = 1,96 \quad (45)$$

$$S[\bar{X}] = \frac{5}{\sqrt{150}} \quad (46)$$

$$KI = 3,5 \pm (1,96 * 0,408) = [2,7 \bullet 4,3] \quad (47)$$

Av de 150 respondentene svarte 67 av de hadde lest minst en bok fra deres forlag.

c) Lag et 95% konfidensintervall for  $\hat{p}$  som estimerer andelen i populasjonen som har lest minst en av deres bøker.

$$\hat{p} = \frac{67}{150} = 0,447 \quad (48)$$

$$KI[\hat{p}_{95}] = 0,447 \pm 1,96 * \sqrt{\frac{0,447 * 0,553}{150}} \quad (49)$$

$$= 0,447 \pm 1,96 * 0,041 = [0,36664 \bullet 0,52736] \quad (50)$$

## 7

I en diskusjon om den optimale lengden på en bok kommer dere inn på lesehastigheter. En undersøkelse viser at gjennomsnittlig antall ord per minutt ( $\mu$ ) er 230 med et standardavvik ( $\sigma$ ) på 37.

- Hva er sannsynligheten for at en tilfeldig valgt person leser mindre enn 200 ord per minutt?
- Dersom du skal oppgi et intervall som favner om 90% av befolkningen, hva er øvre og nedre grense for dette intervallet, i antall ord per minutt?
- Hva er sannsynligheten for at en tilfeldig valgt person leser mellom 210 og 250 ord per minutt?
- Du setter en tilfeldig valgt person til å lese i 60 minutter. Teksten han skal lese er 14000 ord lang. Hva er sannsynligheten av for at personen rekker å lese denne ferdig?

$$P(Z < 200) = G\left(\frac{200 - 230}{37}\right) \quad (51)$$

$$= G(-0,81) = 1 - G(0,81) \quad (52)$$

$$= 1 - 0,7910 = 0,209 \quad (53)$$

$$Z_{0,1/2} = 1,6449 \quad (54)$$

$$KI[90\%] = 230 \pm 1,6449 * 37 = [169,139 \bullet 290,861] \quad (55)$$

$$P(210 < Z < 250) = P(Z < 250) - P(Z < 210) \quad (56)$$

$$= G\left(\frac{250 - 230}{37}\right) - G\left(\frac{210 - 230}{37}\right) \quad (57)$$

$$= G(0,54) - G(-0,54) = G(0,54) - (1 - G(0,54)) \quad (58)$$

$$= 0,7054 - (1 - 0,7054) = 0,7054 - 0,2946 \quad (59)$$

$$= 0,410 \quad (60)$$

$$P(S > 14000) = 1 - P(S < 14000) = \quad (61)$$

$$= 1 - G\left(\frac{14000 - 60 * 230}{\sqrt{60} * 37}\right) \quad (62)$$

$$= 1 - G\left(\frac{200}{286,6}\right) = 1 - G(0,7) \quad (63)$$

$$= 1 - 0,7580 = 0,242 \quad (64)$$

## 8

Dere har tidligere antatt at minst 50% av befolkningen årlig leser en av deres bøker. Nå frykter ledelsen at denne andelen er lavere enn 50%, og du blir bedt om å undersøke dette. Grunnlaget er den siste spørreundersøkelsen, hvor 67 av 150 svarte at de hadde lest en av deres bøker de siste året.

- Foreta en hypotesetest for å avgjøre om andelen som leser minst en av deres bøker har minket. Sett opp  $H_0$ ,  $H_A$  og bruk 95% sikkerhetsnivå.
- Beregn p-verdien til hypotesetesten.

$$H_0 : p \geq 0,5, \quad H_A : p < 0,5 \quad (65)$$

$$Z = \frac{67 - 150 * 0,5}{\sqrt{150 * 0,5 * 0,5}} = \frac{-8}{6,124} = -1,3 \quad (66)$$

$$Z_{0,05} = 1,6449 \quad (67)$$

$$\text{Forkast hvis } Z < -Z_\alpha : -1,3 \not< -1,6449 \quad (68)$$

$$\text{Behold } H_0 \quad (69)$$

$$p = P(Z < -1,3) = G(-1,3) = 1 - G(1,3) \quad (70)$$

$$= 1 - 0,9032 = 0,0968 \quad (71)$$

---