STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik

MT1011, MT102U, MT103U
TENTAMEN
12 januari 2022

Tentamenslösningar – Sannolikhetslära och statistik för lärare

12 januari 2022 kl. 8–13

Examinator: Gudrun Brattström, gudrun@math.su.se

Uppgift 1

a) Variationskoefficienten är kvoten mellan standardavvikelsen och medelvärdet, så den blir $\frac{52\,023}{201\,835} = 0.258$. Vi kan också uttrycka variationskoefficienten i procent: 25.8%.

b) Motsvarande medelinkomst och standardavvikelse uttryckt i euro är 9.33 gånger mindre: $\frac{201\,835}{9.33}€ = 21\,632.90€$ respektive $\frac{52\,023}{9.33}€ = 5\,575.88€$. Variationskoefficienten förändras inte: den förblir 25.8%.

c) Ordna data efter storlek:

```
7.9  8.2  8.6  8.6  9.3  10.1  10.9  12.3  12.9  15.0
```

Medianen är medelvärdet av de två understrukna värdena nedan. Kvartilerna är understrukna dubbelt.

$$7.9 \quad 8.2 \quad \underline{\underline{8.6}} \quad 8.6 \quad \underline{9.3 \quad 10.1} \quad 10.9 \quad \underline{\underline{12.3}} \quad 12.9 \quad 15.0$$

Så $Q_1 = 8.6$, $Q_3 = 12.3$ och $M = Q_2 = \frac{8.2+10.1}{2} = 9.7$.

Om man exkluderar de båda värden som ingick i beräkningen av medianen, så får man istället $Q_1 = \frac{8.2+8.6}{2} = 8.4$ och $Q_3 = \frac{12.3+12.9}{2} = 12.6$.

Uppgift 2

a) Försöksupplägget är att betrakta som stickprov i par eftersom försöken utförs två och två på samma gård, och man kan anta att odlingsförutsättningarna är mera likartade på en och samma gård än de är om man jämför två olika gårdar. De båda försöken på gård nummer i hör ihop.

b) Vi bildar därför differenserna:

Lantbrukare	1	2	3	4	5	6
Medel A	115.8	123.0	122.8	209.8	226.3	150.3
Medel B	121.2	122.1	128.0	214.9	231.2	152.0
Differens	5.4	-0.9	5.2	5.1	4.9	1.7

Om vi kallar de observerade differenserna för d_i $(i = 1, 2, \ldots, 6)$, så har vi $\overline{d} = 3.57$ och $s_d = 2.59$.

Ett 95%-igt konfidensintervall för skillnaden beräknas som $\overline{d} \pm t_{0.025}(5)s_d/\sqrt{6}$ $= 3.57 \pm 2.571 \cdot 2.59/\sqrt{6} = 3.57 \pm 2.72$, eller $(0.85, 6.29)$.

c) Konfidensintervallet i **b)** innehåller inte värdet 0, så vi förkastar nollhypotesen och drar slutsatsen att det *är* skillnad mellan gödningsmedlen.

Uppgift 3

Antalet blåa sockor X i packningen är hypergeometriskt fördelat, $X \sim Hyp(N, n, p)$ med parametrarna $N = 8 + 12 = 20$, $n = 4$ och $p = 8/20 = 0.4$.

a) Väntevärdet för antalet blåa sockor är $np = 4 \cdot 0.4 = 1.6$.

b) Variansen för antalet blåa sockor är enligt formelsamlingen

$$np(1-p)\tfrac{N-n}{N-1} = 4 \cdot 0.4 \cdot 0.6 \cdot \tfrac{20-4}{20-1} = 0.8084\,,$$

så standardavvikelsen blir $\sqrt{0.8084} = 0.899$. (5 p)

c) Om vi betecknar antalet blåa sockor med X så gäller

$$\begin{aligned}
P(\text{Alla samma färg}) &= P(\text{Alla blå}) + P(\text{Alla röda}) = P(X=4) + P(X=0) \\
&= \frac{\binom{8}{4}\binom{12}{0}}{\binom{20}{4}} + \frac{\binom{8}{0}\binom{12}{4}}{\binom{20}{4}} = \frac{\frac{8\cdot7\cdot6\cdot5}{1\cdot2\cdot3\cdot4}\cdot 1 + 1 \cdot \frac{12\cdot11\cdot10\cdot9}{1\cdot2\cdot3\cdot4}}{\frac{20\cdot19\cdot18\cdot17}{1\cdot2\cdot3\cdot4}} \\
&= \frac{8\cdot7\cdot6\cdot5 + 12\cdot11\cdot10\cdot9}{20\cdot19\cdot18\cdot17} = \frac{113}{969} = 0.117\,.
\end{aligned}$$

Uppgift 4

a) Låt X vara komponentens livslängd, så $X \sim \text{Exp}(1500)$. En primitiv funktion till tätheten $f_X(x) = \frac{1}{1500}e^{-x/1500}$, är $F(x) = -e^{-x/1500} + C$. Eftersom en fördelningsfunktion antar värden mellan 0 och 1, så måste $C = 1$ och $F_X(x) = 1 - e^{-x/1500}$. Alltså är

$$P(X > 1350) = 1 - P(X < 1350) = 1 - \big(1 - e^{-1350/1500}\big) = e^{-0.9} = 0.4066\,.$$

b) Vi har

$$\begin{aligned} P(X > 1000 + 1350|X > 1000) &= \frac{P(X > 2350 \cap X > 1000)}{P(X > 1000)} = \frac{P(X > 2350)}{P(X > 1000)} \\ &= \frac{e^{-2350/1500}}{e^{-1000/1500}} = e^{-1350/1500} = e^{-0.9} = 0.4066\,. \end{aligned}$$

Eller också kan vi använda exponentialfördelningens minneslöshet för att direkt dra slutsatsen att sannolikheten är den samma som i **a)**. Det spelar ingen roll vad som har hänt tidigare, komponenten är i varje ögonblick som ny – tills den går sönder.

c) Sannolikheten för att en enskild komponent ska fungera efter 1350 timmar är $e^{-0.9} = 0.4066$. Antalet komponenter Y som fungerar efter 1350 timmar är $\sim \text{Bin}(6, e^{-0.9})$. Eftersom $6/2 = 3$ söker vi

$$\begin{aligned} P(Y \geq 3) &= 1 - P(Y \leq 2) = 1 - (P(Y = 0) + P(Y = 1) + P(Y = 2)) \\ &= 1 - \big(\tbinom{6}{0}(1-e^{-0.9})^6 + \tbinom{6}{1}(1-e^{-0.9})^5 e^{-0.9} + \tbinom{6}{2}(1-e^{-0.9})^4 (e^{-0.9})^2\big) \\ &= 1 - \big(1 \cdot 0.0437 + 6 \cdot 0.0736 \cdot 0.4066 + \tfrac{6\cdot 5}{1\cdot 2} \cdot 0.1240 \cdot 0.1653\big) \\ &= 1 - 0.5307 = 0.4693 \approx 0.47\,. \end{aligned}$$

Om man istället avrundar $e^{-0.9} = 0.4066$ till 0.4 och använder tabellen över kumulativa binomialsannolikheter (Tabell 1), så läser man av, för $p = 0.4$, $n = 6$ och $x = 2$, den kumulativa sannolikheten 0.5443, det vill säga vi får $P(Y \geq 3) = 1 - 0.5443 = 0.4557 \approx 0.46$.

Uppgift 5

a) Beteckna skaldiametern hos en slumpmässigt vald foraminifer ur populationen med X. Vi har alltså $X \sim N(0.625, 0.101)$. Då gäller

$$\begin{aligned} P(X > 0.7) &= P\left(\frac{X - 0.625}{0.101} > \frac{0.7 - 0.625}{0.101}\right) = 1 - P\left(\frac{X - 0.625}{0.101} < \frac{0.7 - 0.625}{0.101}\right) \\ &= 1 - \Phi\left(\frac{0.7 - 0.625}{0.101}\right) = 1 - \Phi(0.74) = 1 - 0.7704 = 0.2296\,, \end{aligned}$$

där $\Phi(x)$ är standardnormalfördelningens fördelningsfunktion.

b) Om $\overline{X}$ betecknar medelvärdet, så gäller $\overline{X} \sim N(0.625, 0.101/\sqrt{3})$, och därför

$$\begin{aligned} P(\overline{X} > 0.7) &= P\left(\frac{\overline{X} - 0.625}{0.101/\sqrt{3}} > \frac{0.7 - 0.625}{0.101/\sqrt{3}}\right) = 1 - P\left(\frac{\overline{X} - 0.625}{0.101/\sqrt{3}} < \frac{0.7 - 0.625}{0.101/\sqrt{3}}\right) \\ &= 1 - \Phi\left(\frac{0.7 - 0.625}{0.101/\sqrt{3}}\right) = 1 - \Phi(1.29) = 1 - 0.9015 = 0.0985\,. \end{aligned}$$

c) Om den största foraminiferen av de tre *inte* har en skaldiameter överstigande 0.7 mm, så betyder det att ingen av dem har det, och omvänt: om ingen av de tre har skaldiameter > 0.7 mm så kan inte den största ha det. Den sannolikhet som ska beräknas är därför

$$1 - P(\text{alla tre har skaldiameter} < 0.7) = 1 - P(X < 0.7)^3 = 1 - 0.7704^3 = 0.5428\,.$$

Uppgift 6

a) Låt X vara antalet gånger du blir attackerad av kråkan. Den sökta sannolikheten är $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1.5} = 1 - 0.2231 = 0.7769$, eller ca 78%.

b) Detta betyder att Poissonintensiteten fördubblas: nu är $\lambda = 3$. Med $X \sim Po(3)$ är sannolikheten att bli attackerad minst en gång $= P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-3} = 1 - 0.0498 = 0.9502$, eller ca 95%.

c) Beteckna den sökta tidrymden med u. Under denna tid är X = antalet attacker från kråkan Poissonfördelat med parameter $\lambda = 1.5u$. Vi vill att $P(X \leq 5) = F_X(5) = 0.8$. I Tabell 2 letar vi upp det λ-värde som ger ett värde så nära 0.8 som möjligt för $F_X(5)$. Vi finner att för $\lambda = 3.9$ är $F_X(5) = 0.8006$. Vi löser ekvationen $1.5u = 3.9$ och finner att $u = 2.6$ timmar. Du kan alltså vistas i parken i 2 timmar och 36 minuter och vara 80% säker på att inte bli attackerad mer än fem gånger av kråkan.