Tentamen Statistische methoden
MST-STM
1 juli 2010, 9:00–12:00

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt—na afloop inleveren alstublieft. Normering: De meerkeuzevragen tellen voor één derde en de open vragen voor twee derde van het cijfer. Bij de open vragen telt elk (vraag)onderdeel even zwaar.

Meerkeuzevragen

Toelichting: In het algemeen zijn niet altijd vijf van de zes alternatieven 100% fout, het juiste antwoord is het meest volledige antwoord. Maak op het bijgeleverde antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Doorstrepen van een fout antwoord heeft geen zin: u moet het òf uitgummen, òf verwijderen met correctievloeistof òf een nieuw formulier invullen. Vergeet niet uw studienummer in te vullen èn aan te strepen.

1. Drie vrienden gaan een avondje naar het casino. Ze spelen daar een spel met kans 1/4 om te winnen. Nadat ze alle drie dit spel drie keer hebben gespeeld, wat is de kans dat precies twee van de drie vrienden geen enkele keer het spel hebben gewonnen?

a. 0.052 **b.** 0.103 **c.** 0.206 **d.** 0.309 **e.** 0.412 **f.** 0.515

2. Zij $X_1, X_2, \ldots, X_n$ een steekproef van positieve getallen uit een verdeling met verwachting λ en variantie λ, waarbij $\lambda > 0$ een onbekende parameter is. Een schatter voor λ wordt gegeven door $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. De onzuiverheid (bias), en MSE (mean square error) van deze schatter worden gegeven door:

a. bias $= 0$, MSE $= \lambda^2/n^2$
b. bias $= 1/(n\lambda)$, MSE $= 2/(n\lambda^2)$
c. bias $= \lambda$, MSE $= 2/(n\lambda) + \lambda^2$
d. bias $= 0$, MSE $= \lambda/n$
e. bias $= 1/(n\lambda)$, MSE $= \lambda/n$
f. bias $= \lambda$, MSE $= \lambda^2/n^2$

3. Een fabrikant van verf wenst de gemiddelde droogtijd van een nieuwe type muurverf te bepalen. Voor 12 testmuren van dezelfde oppervlakte vond hij een gemiddelde droogtijd van 66.3 minuten en een steekproefvariantie van 8.4. De 12 gemeten droogtijden vat men op als een realisatie van een steekproef uit een $N(\mu, \sigma^2)$ verdeling. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

a. (61.944 , 70.655) **b.** (61.978 , 70.621) **c.** (64.730 , 67.869)
d. (64.742 , 67.857) **e.** (64.797 , 67.803) **f.** (64.809 , 67.790)

4. Serrano ham wordt in een winkel verkocht in verpakkingen van 100 gram. De manager vreest echter dat het personeel systematisch iets te veel ham in de verpakkingen stopt. Hij meet de hoeveelheid Serrano ham in $n = 12$ verpakkingen. Noem de uitkomsten hiervan $x_1, \ldots, x_{12}$. Gegeven is dat

$$\bar{x}_n = 101.03\,g \quad \text{en} \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = 4.00\,g^2.$$

De manager gaat er vanuit dat het gewicht van de ham in een verpakking verdeeld is volgens $N(\mu, \sigma^2)$. We toetsen m.b.v. de t-toets de nul-hypothese $H_0 : \mu = 100$ tegen het alternatief $H_1 : \mu > 100$, bij een significantieniveau van 5%. Noem p de p-waarde van deze toets. Welke van de volgende zes conclusies is de juiste?

a. $0.025 < p < 0.05$, verwerp H_0.
b. $0.025 < p < 0.05$, verwerp H_0 niet.
c. $0.05 < p < 0.10$, verwerp H_0.
d. $0.05 < p < 0.10$, verwerp H_0 niet.
e. $p > 0.10$, verwerp H_0.
f. $p > 0.10$, verwerp H_0 niet.

5. Een docent ziet dat een bepaalde moeilijke vraag tijdens een multiple choice tentamen met 6 keuze mogelijkheden, door relatief weinig studenten goed is beantwoord. De docent wil graag weten hoeveel studenten het antwoord echt wisten. Hij gebruikt het volgende model: een student heeft kans p om het goede antwoord te *weten*. Als de student het antwoord niet weet, gokt hij of zij volledig willekeurig een van de zes antwoorden. Noem q de kans dat een student het juiste antwoord *geeft*. Welke van de volgende uitspraken is juist volgens het model van de docent?

a. $p = \frac{5}{6}q$ **b.** $p = \frac{1}{6}q + \frac{1}{6}$ **c.** $p = q - \frac{1}{6}$
d. $p = q$ **e.** $p = \frac{6}{5}(q - \frac{1}{6})$ **f.** $p = \frac{6}{5}q - \frac{1}{6}$

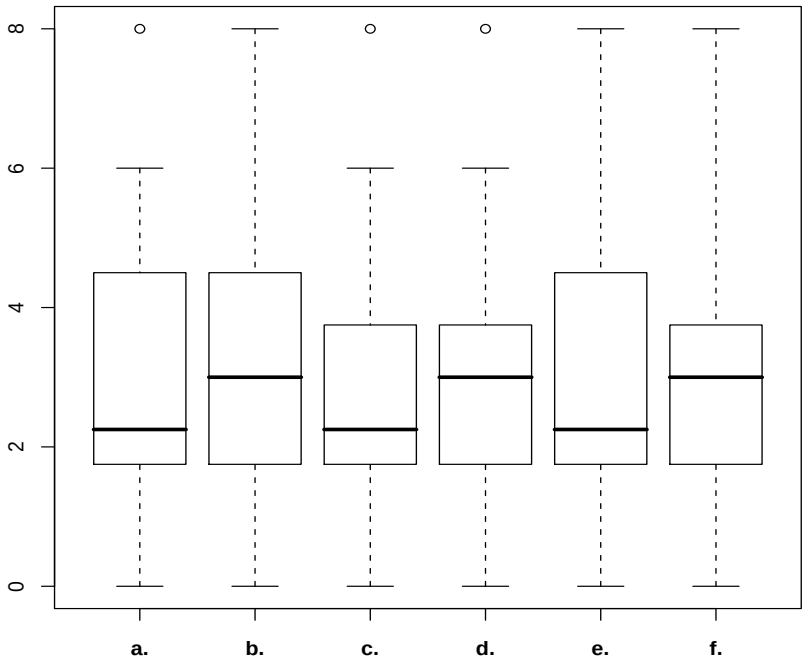
6. Het gewicht van pindakaas in een pot is normaal verdeeld, met verwachting 120g en standaarddeviatie 3g. De pot zelf heeft een gewicht dat ook normaal verdeeld is, met verwachting 60g en standaarddeviatie 2g. Het gewicht van de pot zelf is onafhankelijk van het gewicht van de pindakaas in de pot. Wat is de kans dat 10 potten pindakaas in totaal meer wegen dan 1820g? U mag gebruiken dat de som van onafhankelijke normale verdelingen, weer normaal verdeeld is.

a. 0.0020 **b.** 0.0397 **c.** 0.1038 **d.** 0.3446 **e.** 0.9000 **f.** 0.9980

7. Gegeven is de dataset

0.00 1.5 1.75 1.75 2.00 3.00 3.25 3.75 3.75 6.00 8.00

Welke van de zes onderstaande boxplots hoort bij deze dataset?



8. Een afvulmachine van een frisdrankfabriek vult flessen met een hoeveelheid drank die normaal verdeeld is met verwachtingswaarde μ en standaardafwijking $\sigma = 5$ ml. Van 25 flessen is de gemiddelde inhoud $\bar{x}_n = 1502$ ml. Het 95% betrouwbaarheidsinterval voor de verwachte inhoud van een fles wordt gegeven door

a. $1502 \pm z_{0.025} \cdot 5/\sqrt{24}$
b. $1502 \pm t_{0.05,24} \cdot 5/\sqrt{24}$
c. $1502 \pm t_{0.05,24}$
d. $1502 \pm t_{0.025,24}$
e. $1502 \pm z_{0.05}$
f. $1502 \pm z_{0.025}$

9. We toetsen een hypothese H_0 op significantieniveau 0.05. Dit betekent dat:

a. als H_0 onjuist is, de kans op verwerpen van H_0 minstens 0.95 is.
b. als H_1 onjuist is, de kans op verwerpen van H_1 minstens 0.95 is.
c. als H_0 juist is, de kans op niet verwerpen van H_0 minstens 0.95 is.
d. als H_1 juist is, de kans op niet verwerpen van H_1 minstens 0.95 is.
e. als H_0 onjuist is, de kans op verwerpen van H_0 minstens 0.05 is.
f. als H_0 juist is, de kans op niet verwerpen van H_0 minstens 0.05 is.

10. We beschikken over een dataverzameling $x_1, x_2, \ldots, x_n$ met gemiddelde $\bar{x}_n$. De dataverzameling is een realisatie van een steekproef $X_1, X_2, \ldots, X_n$ uit een $Exp(\lambda)$ verdeling. De steekproefgrootheid $1/\bar{X}_n$ is een schatter voor λ. Men wil de kansverdeling van de stochast $T_n = 1/\bar{X}_n - \lambda$ benaderen door middel van een bootstrapsimulatie. Met wat voor soort bootstrapprocedure hebben we hier te maken en wat is de bootstrapversie van T_n?

a. een parametrische bootstrap met $T_n^* = \bar{X}_n^* - 1/\bar{x}_n$.
b. een parametrische bootstrap met $T_n^* = 1/\bar{X}_n^* - \bar{x}_n$.
c. een parametrische bootstrap met $T_n^* = 1/\bar{X}_n^* - 1/\bar{x}_n$.
d. een empirische bootstrap met $T_n^* = \bar{X}_n^* - 1/\bar{x}_n$.
e. een empirische bootstrap met $T_n^* = 1/\bar{X}_n^* - \bar{x}_n$.
f. een empirische bootstrap met $T_n^* = 1/\bar{X}_n^* - 1/\bar{x}_n$.

Open vragen

Toelichting: Een antwoord alleen is *niet* voldoende: er dient een berekening, toelichting en/of motivatie aanwezig te zijn. Dit alles goed leesbaar en in goed Nederlands.

1. Gegeven zijn metingen aan deeltjesdiameters x_1, x_2, …, x_n en waarvan we aannemen dat ze afkomstig zijn van een steekproef aan een geschaalde $Par(\alpha)$ verdeling, dat wil zeggen, met dichtheid

$$f(x) = \frac{\alpha}{m}(x/m)^{-(\alpha+1)} \quad \text{voor } x > m,$$

en $f(x) = 0$ elders. Er zijn twee parameters: $\alpha > 0$ en $m > 0$.

a. Bepaal de bijbehorende verdelingsfunctie en geef de formule voor de mediaan.

b. Bepaal de likelihood functie $L(\alpha, m)$ voor deze dataset en beredeneer dat de maximum likelihood schatter voor m gelijk is aan $\hat{m} = \min(x_1, x_2, \ldots, x_n)$. Laat vervolgens zien dat die voor α gelijk is aan $\dfrac{n}{\sum_{i=1}^{n} \ln(x_i/\hat{m})}$.

c. Leg uit wat zuiverheid inhoudt en maak aannemelijk dat geen van beide schatters zuiver is.

2. In een laboratorium moeten 1000 monsters getest worden op aanwezigheid van zware metalen. Eén mogelijkheid is het uitvoeren van 1000 tests. Een mogelijk efficiëntere aanpak is de volgende. Men voegt telkens (gedeelten van) 25 monsters bij elkaar en test het mengsel op sporen. Indien hierin sporen worden aangetroffen test men alle 25 monsters alsnog individueel.

Omdat in de praktijk circa 2 procent van de monsters besmet is, willen we dit schema analyseren onder de volgende modelaanname: de kans dat een monster sporen bevat ('positief' is) is gelijk aan 0.02, onafhankelijk van wel/geen aanwezigheid van sporen in andere monsters.

a. Het aantal tests dat voor een groep van 25 uitgevoerd moet worden is een stochastische variabele, zeg N. Bepaal de kansverdeling van N en ook de verwachting.

Een slimmerik stelt de volgende subtielere aanpak voor: deel de groep van 25 op in vijf subgroepjes van vijf monsters; neem per subgroepje van elk monster een beetje en maak een 'subgroepmengsel'; neem van elk van de subgroepmengsels een beetje en maak, door dit te mengen, een 'overall mengsel;' test nu eerst het 'overall mengsel'; als dat positief is, test dan alle 'subgroepmengsels;' en als een subgroepmengsel positief is, test dan elk van de vijf monsters ook individueel.

b. Definieer Y als het aantal van de vijf subgroepmengsels dat positief test. Geef de kansverdeling van Y en de verwachting.

c. De slimmerik redeneert: als Y van de 5 groepsmengsels positief zijn doe ik in totaal dus $M = 1 + 5 + Y \cdot 5$ tests (alles+groepjes+individueel). Bepaal $\mathrm{E}[M]$.

d. Er zit een denkfout in de redenering hierboven. Identificeer hem en doe vervolgens een zo scherp mogelijke uitspraak over het verwachte aantal tests.

3. De stochasten X en Y hebben de volgende kansdichtheid:

$$f(x,y) = \frac{12}{5}xy(1+y), \quad 0 \le x \le 1, 0 \le y \le 1$$

en buiten dit gebied geldt $f(x,y) = 0$.

a. Bepaal de marginale kansdichtheid van Y.

b. De marginale kansdichtheid van X is $f_X(x) = 2x$ voor $0 \le x \le 1$ en daarbuiten geldt $f_X(x) = 0$. Bepaal $\mathrm{E}[7X - 2]$ en $\mathrm{Var}(7X - 2)$.

c. Gegeven is bovendien dat $\mathrm{E}[Y] = \frac{7}{10}$. Bepaal $\mathrm{Cov}(X, Y)$.

Antwoorden multiple choice:

1 d. De kans dat vriend nummer 1 geen een keer wint, is gelijk aan $p = (1 - \frac{1}{4})^3 = 0.422$. De kans dat precies twee van de drie vrienden geen een keer wint, is dus de binomiale kans

$$\mathrm{P}(Bin(3,p) = 2) = \binom{3}{2} p^2 (1-p)^1 = 0.309.$$

2 d. Bekend is dat $\mathrm{E}\left[\bar{X}_n\right] = \mathrm{E}[X_1] = \lambda$, en $\mathrm{Var}\left(\bar{X}_n\right) = \frac{\mathrm{Var}(X_1)}{n} = \lambda/n$. De schatter is zuiver zodat de bias gelijk is aan 0. Omdat de schatter zuiver is, is de MSE gelijk aan de variantie van de schatter. Het goede antwoord is dus

$$\text{bias} = 0, \quad \text{MSE} = \lambda/n.$$

3 e. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

$$\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}},\, \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right).$$

Hierbij is $t_{n-1,\alpha/2} = t_{11,0.05} = 1.796$ en $s_n = \sqrt{8.4} = 2.898$, zodat het 90% betrouwbaarheidsinterval wordt gegeven door

$$\left(66.3 - 1.796\frac{2.898}{\sqrt{12}},\, 66.3 + 1.796\frac{2.898}{\sqrt{12}}\right) = (64.797,\, 67.803).$$

4 d. De toetsingsgrootheid wordt gegeven door

$$T = \frac{\bar{x}_n - 100}{s_n/\sqrt{n}} = 1.784.$$

Er geldt dat $t_{11,0.10} < T < t_{11,0.05}$ (merk op dat T een t-verdeling heeft met $n - 1 = 11$ vrijheidsgraden!). Aangezien alleen grote waarden van T wijzen op de alternatieve hypothese H_1, geldt dus dat $0.05 < p < 0.10$. Aangezien p groter is dan het significantieniveau, verwerpen we H_0 niet.

5 e. Er geldt dat een student een vraag goed beantwoordt als hij het weet, of als hij het niet weet en goed gokt. Dus

$$q = p + (1-p)\cdot\frac{1}{6} \quad \Leftrightarrow \quad p = \frac{6}{5}\left(q - \frac{1}{6}\right).$$

6 b. Het totale gewicht van een pot is normaal verdeeld met verwachting $120 + 60 = 180$. De variantie van het totale gewicht is de som van de varianties: $3^2 + 2^2 = 13$. Het totale gewicht W van tien potten is weer normaal verdeeld en heeft verwachting $10 \cdot 180 = 1800$, en variantie $10 \cdot 13 = 130$. Er geldt dus

$$\mathrm{P}(W > 1820) = \mathrm{P}\left(\frac{W - 1800}{\sqrt{130}} > \frac{1820 - 1800}{\sqrt{130}}\right) = \mathrm{P}(Z > 1.75) = 0.0397.$$

7 d. We bereken eerst de mediaan: het 6de getal (van de geordende dataset met 11 punten) is 3, dus de mediaan is 3. Dan zijn alleen **b**, **d** en **f** nog over. Vervolgens berekenen we het derde kwartiel: $\frac{3}{4} \cdot 11 = 8.25$ (volgens het boek: $\frac{3}{4} \cdot (11 + 1) = 9$), en het 8ste en 9de getal zijn beide 3.75, dus het derde kwartiel is 3.75. Dan blijven alleen **d** en **f** over. Tot slot berekenen we de Inter Quartile Range: het onderste kwartiel is 1.75 (de waarde van het derde en vierde getal), dus de IQR = 2. De omhooggaande "whisker"heeft maximale lengte $1.5 \cdot \text{IQR} = 3$, en $6 < 3.75 + 3 < 8$. Dit betekent dat de whisker tot aan 6 gaat. Dit leidt tot antwoord **d**.

8 f. De variantie $\sigma^2 = 25$ is bekend, en we weten dat voor de echte onbekende verwachting μ geldt:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ heeft een } N(0,1) \text{ verdeling.}$$

Hieruit volgt dat

$$\mathrm{P}\big(\bar{X}_n - z_{0.025}\sigma/\sqrt{n} \le \mu \le \bar{X}_n + z_{0.025}\sigma/\sqrt{n}\big) = \mathrm{P}\big(\bar{X}_n - z_{0.025} \le \mu \le \bar{X}_n + z_{0.025}\big) = 0.95.$$

9 c. Het significantieniveau is de maximaal toelaatbare kans op een type I fout, waarbij een type I fout overeenkomt met verwerpen van H_0, terwijl deze in werkelijkheid juist is. Als het significantieniveau 0.05 is, dan betekent dit dus dat in het geval dat H_0 juist is, de kans op verwerpen hoogstens 0.05 en dus is de kans op niet verwerpen minstens 0.95.

10 c. De kansverdeling van de X_i's is van een parametrisch type: een $Exp(\lambda)$ verdeling. Zodoende hebben we te maken met een parametrische bootstrap. De parameter λ schatten we door $\hat{\lambda} = 1/\bar{x}_n$, en dus is de bootstrap versie van T_n gelijk aan de stochast $T_n^* = 1/\bar{X}_n^* - \hat{\lambda} = 1/\bar{X}_n^* - 1/\bar{x}_n$.

Antwoorden open vragen:

1a $F(x) = 0$ voor $x \le m$ en $F(x) = 1 - (x/m)^{-\alpha}$ voor $x > m$. Behalve door zelf te primitiveren kan dit antwoord afgeleid worden uit de formules voor de $Par(\alpha)$ op het formuleblad. Voor de mediaan lossen we op $F(q) = 0.5$ en vinden $(q/m)^{\alpha} = 0.5$ ofwel $q = 2^{1/\alpha}\, m$.

1b We schrijven $f(x) = \alpha \cdot m^{\alpha} \cdot x^{-(\alpha+1)}$, voor $x \ge m$. Het voorbeeld op pagina 318, 319, en ook opgaven 21.6 en 21.9 laten zien dat $L(\alpha, m) = 0$ als $m > x_i$ voor een of andere x_i. Voor $m \le min(x_1, \ldots, x_n)$ geldt

$$L(\alpha, m) = f(x_1) \cdot \ldots \cdot f(x_n) = \alpha^n \cdot m^{n\alpha} \cdot (x_1 \cdots x_n)^{-(\alpha+1)}\,.$$

Voor vaste α gedraagt L zich als een constante maal $m^{n\alpha}$, een positieve macht van m, dus stijgend, maar na $min(x_1, \ldots, x_n)$ opeens nul. Het maximum ligt dus op de rand.
De tweede parameter volgt door $\ln(L) = n\ln(\alpha) + n\alpha\ln(m) - (\alpha+1)\sum_{i=1}^{n}\ln(x_i)$ te differentiëren naar α en gelijk aan 0 te stellen, hetgeen $\frac{n}{\alpha} + n\ln(m) - \sum_{i=1}^{n}\ln(x_i) = 0$ oplevert, waaruit het antwoord volgt. Merk op dat de tweede afgeleide naar α negatief is: het is inderdaad een maximum.

1c Voor de theorie zie Hoofdstuk 19. Het minimum van de metingen zal nooit precies op de ondergrens m liggen maar altijd er boven. Ergo, er is een positieve bias voor m en omdat α daarvan afhangt zal deze waarschijnlijk ook onzuiver zijn (geavanceerde analyse—buiten ons bestek—laat zien dat deze bias ook positief is).

2a Per groep van 25 wordt één test uitgevoerd. De kans dat geen sporen worden gedetecteerd, is P(geen van de 25 monsters bevat sporen) $= (1 - 0.02)^{25} = 0.98^{25} = 0.6035$. In het complementaire geval worden er 25 extra tests uitgevoerd. De stochastische variabele N kan de waarden 1 en 26 aannemen, met kansen respectievelijk 0.6035 en 0.3965. Ergo, $\mathrm{E}[N] = 1 + 25\cdot(1 - 0.98^{25}) = 10.91$.

2b Per groep van 5 wordt één test uitgevoerd. De vijf groepen testen elk positief met dezelfde kans en onafhankelijk van elkaar: Y heeft dus een binomiale verdeling (zie Hoofdstuk 4). Nu de parameters nog. De kans dat geen sporen worden gedetecteerd, is P(geen van de 5 monsters bevat sporen) $= (1-0.02)^5 = 0.98^5 = 0.9039$. Ergo, de kans dat sporen worden gedetecteerd is $1 - 0.98^5 = 0.0961$.
De verdeling van Y is dus $Bin(5, 0.0961)$ en daaruit volgt direct: $\mathrm{E}[Y] = 5 \times 0.0961 = 0.4804$.

2c Lineariteit van de verwachting: $\mathrm{E}[M] = 6 + 5 \cdot \mathrm{E}[Y] = 8.402$.

2d Als $Y = 0$ dan is geen van de monsters besmet en volstaat de eerste test; er zijn er dus geen 6 nodig, zoals de formule voor M in dit geval geeft. Voor $Y > 0$ geeft de formule de juiste uitkomst. Eerste conclusie: het verwachte aantal tests is dus *kleiner dan* $\mathrm{E}[M]$.
Hoeveel kleiner? In plaats van $\mathrm{E}[M] = \mathrm{E}[6 + 5Y] = \sum_{k=0}^{5}(6 + 5 \cdot k)\mathrm{P}(Y = k)$ moeten we in deze uitdrukking voor $k = 0$ de 6 vervangen door een 1, de rest blijft ongewijzigd. Het verschil is $5 \cdot \mathrm{P}(Y = 0) = 5 \cdot 0.98^{25} = 3.017$. Het feitelijke verwachte aantal tests is $8.402 - 3.017 = 5.385$.

Een algemenere probleemstelling vraagt naar de optimale opsplitsing van een groep van n in g groepjes van m. Voor integer m en g is dan het totaal verwachte aantal tests gelijk aan $g(1 + m[1 - (1 - p)^m])$ en voor vaste n ligt de optimale waarde van m in de buurt van $1/\sqrt{p}$. Bij $p = 0.02$ komt hier 7.07 uit en voor $n = 25$ is de optimale verdeling waarschijnlijk $6 + 6 + 6 + 7$ of $8 + 8 + 9$.

3a We bepalen de marginale kansdichtheid door de andere variabele eruit te integreren:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}x = \int_0^1 \frac{12}{5}xy(1 + y)\,\mathrm{d}x = \frac{6}{5}y(1 + y) \quad 0 \le y \le 1;$$

buiten dit gebied geldt $f_Y(y) = 0$.
Een alternatieve weg naar dit antwoord loopt via de simultane verdelingsfunctie (vrij omslachtig, maar niet fout).

3b We weten dat $\mathrm{E}[7X - 2] = 7\mathrm{E}[X] - 2$ en $\mathrm{E}[X]$ is uit de definitie te bepalen:

$$\mathrm{E}[X] = \int_0^1 x\,2x\,\mathrm{d}x = \frac{2}{3}.$$

Er volgt: $\mathrm{E}[7X - 2] = 7 \cdot \frac{2}{3} - 2 = \frac{8}{3} \approx 2.667$.
Voor $\mathrm{Var}(7X - 2)$ bepalen we eerst $\mathrm{Var}(X)$ via $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$ en vervolgens gebruiken we $\mathrm{Var}(7X - 2) = 49\,\mathrm{Var}(X)$:

$$\mathrm{E}\left[X^2\right] = \int_0^1 x^2\,2x\,\mathrm{d}x = \frac{1}{2}.$$

$$\mathrm{Var}(X) = \frac{1}{2} - \frac{4}{9} = \frac{1}{18} \quad \text{en} \quad \mathrm{Var}(7X - 2) = 49 \cdot \mathrm{Var}(X) = \frac{49}{18} \approx 2.722.$$

3c We zien dat $f(x, y)$ het produkt is van zijn marginale kansdichtheden, dus X en Y zijn onafhankelijk. Derhalve volgt zonder berekening dat $\mathrm{Cov}(X, Y) = 0$.
Heb je dit niet gezien, dan moet je

$$\mathrm{E}[XY] = \int_{x=0}^1 \int_{y=0}^1 xyf(x, y)\mathrm{d}y\,\mathrm{d}x = \int_{x=0}^1 \int_{y=0}^1 xy\frac{12}{5}x\,y(1 + y)\,\mathrm{d}y\,\mathrm{d}x$$

bepalen. De dubbelintegraal is te schrijven als een produkt van enkelvoudige integralen, en het antwoord $7/15 \approx 0.467$ is snel gevonden. De covariantie volgt dan uit $\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$, met 0 als antwoord.