
Skriftlige prøveopgaver i dataanalyse – sæt 2

Fysik 2 & Nano 4

Kursusholder: Ege Rubak

Eksamensdato: Tirsdag den 11. Maj 2010

Bemærk: Til eksamen vil der typisk indgå 2-3 af sådanne dataanalyse opgaver sammen med 2-3 differentiaalligningsopgaver.

Ved bedømmelsen vil der blive lagt vægt på såvel korrekt metode og korrekt svar derfor skal metoden klart fremgå af besvarelsen – held og lykke.

Opgave 1. En stokastisk variabel X har middelværdi 10 og varians 50.

1. Beregn middelværdi og varians af den stokastiske variabel $Y = 10 + 5X$.
 $E(Y) = 60$
2. Find middelværdien af $V = (X - 10)^2$ og $Z = X^2$.
 $E(V) = 50$, $E(Z) = 150$

Opgave 2. En grundejerforening udsender et spørgeskema, hvor man bl.a. spørger, om man er villig til at betale mere for en bedre snerydning. De får 84 svar fordelt efter alder som:

Alder	Nej	Ja	Svarer ikke
20–25	1	0	0
26–35	0	3	1
36–50	6	10	10
51–60	1	7	1
61–70	2	13	6
> 70	4	13	6

1. Beregn de betingede sandsynligheder for svaret er ja givet, at personens alder er hhv. ≤ 50 og > 50 år.
 $P(\text{Ja}|\text{Alder} \leq 50) = 0.42$, $P(\text{Ja}|\text{Alder} > 50) = 0.62$
2. Er hændelserne $A = \{\text{Alder} \leq 50 \text{ år}\}$ og $B = \{\text{Ja}\}$ uafhængige? Begrund dit svar.
De er IKKE uafhængige. Dvs. de er afhængige.

Opgave 3. Når en metaldetektor i en lufthavn signalerer metal ved en passager, er der 25% sandsynlighed for, at signalet skyldes mønter i lommen.

1. En dag bliver 15 passagerer stoppet ved et signal fra metaldetektoren. Hvad er sandsynligheden for, at mindst 3 af disse passagerer har mønter i deres lommer?
Brug binomialfordelingen. $P(X \geq 3) = 1 - P(X \leq 2) = 0.7639$
2. Delspørgsmål 1 fortsat: Er det sandsynligt, at ingen af disse 15 passagerer er stoppet pga. mønter i lommerne? Forklar dit svar udfra sandsynligheden for, at netop denne hændelse optræder.
 $P(X = 0) = 0.0134$, ret usandsynligt.

Opgave 4. Som bekendt er instituttets netværk ofte nede, og op til projektafslutningen beslutter nogle studerende sig for at måle den daglige nedetid i minutter. De måler derfor hver dag i 14 dage antal minutter, nettet dagligt er nede, og opnår flg. resultater:

Dag	1	2	3	4	5	6	7
Nedetid (minutter)	229	295	343	337	282	313	262
Dag	8	9	10	11	12	13	14
Nedetid (minutter)	303	201	374	376	343	406	163

Nedetiderne antages at være uafhængige normalfordelte med ens middelværdi, μ , og ens varians σ^2 .

1. Estimer middelværdien μ og spredningen σ for den daglige nedetid.
 $\bar{x} = 301.93$, $s^2 = 4852.687$, $s = 69.66$.
2. Bestem et 95% konfidensinterval for middelværdien μ .
 $[308.73; 295.13]$.
3. De studerende vil jo gerne, at nettet er nede så lidt som muligt. Test derfor på 5% signifikansniveau, om de kan regne med, at den daglige nedetid er signifikant mindre end 4 timer, altså 240 minutter.
 $H_0 : \mu \geq 240$, $H_1 : \mu < 240$. Kritisk værdi, $t_\alpha(13) = -1.77$ (da testet kun er en-sidedet er det kun negative værdier af teststørrelsen der er kritisk). Teststørrelse fra data, $t_{\text{obs}} = 3.33$. Tydeligvis er 3.33 ikke mindre end -1.77 og vi kan altså ikke afvise nulhypotesen om at nettet er nede i mindst 4 timer. Dvs. den er ikke signifikant mindre end 4 timer.

Opgave 5. Til en julefrokost har 10 studerende målt deres alkoholpromille ved hver at puste to gange i et alkoholmeter. De opnår flg. målinger, hvor også forskellen mellem de to puste er angivet:

Studerende	1. pust	2. pust	Forskel
1	0.9	0.9	0.0
2	1.0	1.8	-0.8
3	1.8	1.8	0.0
4	1.2	1.6	-0.4
5	0.8	0.8	0.0
6	1.0	0.8	0.2
7	0.9	1.0	-0.1
8	1.2	2.1	-0.9
9	2.2	2.0	0.2
10	1.2	1.5	-0.3
\bar{x}	1.22	1.43	-0.21
s^2	0.197	0.260	0.150

Det antages, at de stokastiske variable hørende til promillerne indenfor hhv. 1. pust og 2. pust er uafhængige normalfordelte med ens middelværdi og ens varians. Bemærk, at de to puste for den samme studerende **ikke** er uafhængige.

1. Find et 90% konfidensinterval for forskellen mellem de to puste.
 -0.21 ± 0.224 .
2. Test på 5% signifikansniveau om niveauet ved 1. pust er forskellig fra niveauet ved 2. pust.
Vi kan ikke afvise at de to målinger er ens.

Husk at angive studienummer på alle afleverede ark samt hvor mange sider din besvarelse består af.
