Universitetet i Stavanger

EKSAMEN I: STA100 SANNSYNLIGHETSREGNING OG STATISTIKK 1

VARIGHET: 4 TIMER DATO: 13. MAI 2013 BOKMÅL

TILLATTE HJELPEMIDLER:

KALKULATOR: **HP30S**, **Citizen SR-270X**, **Casio FX82** eller **TI-30**

OPPGAVESETTET BESTÅR AV 4 OPPGAVER PÅ 9 SIDER INKL. VEDLEGG

Oppgave 1

En tilfeldig variabel Y har fordelingen:

y	1	2	3	4
$P(Y=y)$	0.3	0.4	0.2	0.1

a) Tegn histogrammet over fordelingen til Y.
Hva er $P(Y \leq 1)$, og hva er $P(Y > 2)$?
Finn forventning, $\mathrm{E}(Y)$, og varians, $\mathrm{Var}(Y)$, til Y.

b) En ny tilfeldig variabel X er gitt ved at $X = Y + 2$. Finn forventning og varians til X.
Tegn histogrammet over fordelingen til X i en figur sammen med histogrammet over fordelingen til Y. Kommenter kort (forskjeller/likheter).

Vi ser igjen på Y med fordelingen gitt øverst i oppgaven.

c) Hva er $P(Y = 1 | Y \leq 2)$?
Hva er den betingede sannsynligheten for at $Y = 4$ gitt at Y er minst 3?

Oppgave 2

Konsentrasjonen sink i en elv blir målt. Resultatet av en måling betraktes som utfall av en tilfeldig variabel, X, som antas å være normalfordelt med forventning μ og stanfdardavvik σ, $X \sim N(\mu, \sigma)$.

Først skal vi anta at $\mu = 2.5$ og $\sigma = 0.3$ er kjent.

a) Finn $P(X < 1.9)$, og finn $P(X > 2.8)$.
Hva er sannsynligheten for å få et måleresultat mellom 2.2 og 2.8?

b) Anta at resultatene av ulike målinger er uavhengige av hverandre. For fire måleresultat, X_1, X_2, X_3, X_4, finn forventning og standardavvik til gjennomsnittet, $\overline{X} = \frac{1}{4}\sum_{i=1}^{4} X_i$.
Hva er sannsynligheten for gjennomsnittet er større enn 2.8?

Oppgave 3

Hjerteraten (puls; slag pr. minutt) til ni pasienter er målt før (ved tid 0) de fikk en bestemt medisin (enalaprilat) og 30 minutt etter.[1]

Resultatene er gitt i tabellen til høyre. I kollonnen lengst til høyre er differansen, puls ved 30 minutter minus puls ved 0 minutter for hver pasient oppgitt. Gjennomsnitt og empirisk standardavvik for dataene i de tre kollonnene er oppgitt nederst.

Vi betrakter målingene ved tid 0 som utfall av 9 u.i.f. tilfeldige variable $X_1, \ldots, X_9$, der $\mathrm{E}(X_i) = \mu_X$ og $\mathrm{Var}(X_i) = \sigma_X^2$; tilsvarende $Y_1, \ldots, Y_9$, med $\mathrm{E}(Y_i) = \mu_Y$ og $\mathrm{Var}(Y_i) = \sigma_Y^2$ for målingene ved tid 30.

	Tid i minutter		
pasient	0	30	differanse
1	97	92	-5
2	111	106	-5
3	90	86	-4
4	96	78	-18
5	129	124	-5
6	101	98	-3
7	73	68	-5
8	80	75	-5
9	101	106	5
gj.sn.	97.6	92.6	-5
emp.std.	16.4	17.8	5.9

a) Lag et 95% konfidensintervall for forventet hjerterate ved tid 0. Hvilke forutsetninger bygger gyldigheten av dette intervallet på?

b) Er det forskjell i forventet puls ved 0 og 30 minutter? Besvar spørsmålet ved å gjennomføre en toutvalgstest av $H_0 : \mu_X = \mu_Y$ mot $H_1 : \mu_X \neq \mu_Y$.
Gjør klart hva som er teststørrelse, nullfordeling og forkastningsområde.

c) Gjennomfør en test etter parplanen for å besvare spørmålet om det er forskjell i forventet puls ved 0 og 30 minutter. Hva blir konklusjonen nå?
Kommenter eventuelle ulikheter i resultatene i b) og c).
I hvilke siuasjoner bør man bruke parplanen?

Oppgave 4

Et bestemt stoff tilsatt bensinen antas å kunne redusere utslipp av nitrogenoksid (NO_X) fra bilmotorer. For å undersøke dette nærmere, ble mengde tilsetningsstoff, x, og reduksjon i NO_X-utslipp, y, målt for $n = 10$ biler.[2] Dataene er gitt i tabellen under.

x_i	1	1	2	3	4	4	5	6	6	7
y_i	2.1	2.5	3.1	3	3.8	3.2	4.3	3.9	4.4	4.8

Vi vil gjøre en regresjonsanalyse av dataene for å studere forventet reduksjon i NO_X-utslipp som funksjon av x.

a) Lag et spredningsdiagram av dataene og tegn inn omtrentlig regresjonslinjen.
Still opp modellen for enkel regresjonsanalyse. Hva er vanlige antakelser? Gi en fortolkning av stigningstallet i modellen.

[1]Data med utgangspunkt i tabell 12.2 i Altman: *Practical statistics for medical research*.
[2]Data hentet fra tabell 10.2 i Bhattacharyya and Johnson: *Statistical concepts and methods*).

For disse dataene er:

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$:	3.9
$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$:	3.51
$S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$:	40.90
$S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$:	6.85
$S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n}(x_i - \overline{x})y_i$:	15.81

b) Beregn estimatene av konstantledd og stigningstall i regresjonsmodellen.
Det oppgis her at SSE er 0.7376. Hva blir estimatet av residualvariansen (støyvariansen)?
Er det sammenheng mellom mengde tilsetningsstoff og forventet reduksjon i NO_X-utslipp?
Besvar spørsmålet ved hjelp av en passende hypotesetest. Gjør klart hva H_0 og H_1 er.

c) Beregn residualene for de tre første dataparene $(1, 2.1), (1, 2.5), (2, 3.1)$. Dersom du ikke har beregnet estimatene i b), kan du bruke $a = 1.7$ og $b = 0.5$.

Beregn R^2 og forklar hva denne størrelsen viser.