Høgskolen i Østfold

EKSAMEN

Emnekode:	**Emnenavn:**
SFB107111	Metode 1, statistikk deleksamen
Dato:	**Eksamenstid:**
7. mai 2018	4 timer
Hjelpemidler:	**Faglærer:**
Godkjent kalkulator og vedlagt formelsamling m/tabeller	Hans Kristian Bekkevard

Om eksamensoppgaven og poengberegning:

Oppgavesettet består av 15 sider inklusiv denne forsiden, hvorav de 10 siste sidene er formelsamling og tabeller. Kontroller at oppgaven er komplett før du begynner.

Oppgavesettet består av 25 deloppgaver. Alle deloppgavene skal besvares og teller likt ved sensureringen.

Dersom du mener at noe i oppgaven er uklart, ta selv de forutsetninger du mener er nødvendige.

Lykke til.

Sensurfrist: 28. mai 2018

Karakterene er tilgjengelige for studenter på Studentweb www.hiof.no/studentweb

Oppgave 1

Anta at 40 % av alle studenter bor hjemme mens de studerer. Anta videre at 80 % av alle studenter har jobb ved siden av studiene, og at 30 % av alle studenter BÅDE bor hjemme mens de studerer OG har jobb ved siden av studiene.

La H angi begivenheten at en student bor hjemme og la J angi begivenheten at en student har jobb ved siden av studiene.

a) Hva er sannsynligheten for at en student enten bor hjemme eller har jobb ved siden av studiene eller både bor hjemme og har jobb ved siden av?

b) En tilfeldig valgt student har jobb ved siden av studiene. Hva er sannsynligheten for at denne studenten bor hjemme?

Omtrent 8 % av alle menn og 0,64 % av alle kvinner er fargeblinde. Videre kan du legge til grunn at det er 50 % menn og 50 % kvinner i verden.

c) Hva er sannsynligheten for at en tilfeldig person er fargeblind?

d) Blant alle fargeblinde velger vi tilfeldig ut en person. Hva er sannsynligheten for at denne personen er en kvinne?

Oppgave 2

Et varelager inneholder 40 oppblåsbare juletrær. 10 av disse er punktert (ødelagt). Vi tar med oss tre tilfeldige juletrær fra lageret.

a) På hvor mange måter kan 3 juletrær trekkes ut?

b) Hva er sannsynligheten for at alle de 3 juletrærne du trekker ut er i orden (ikke punktert)?

c) Hva er sannsynligheten for at minst et juletre av de 3 du trekker ut er punktert?

Oppgave 3

I en spørreundersøkelse har 100 personer blitt spurt om sin mening om et nytt prosjekt. 46 sier de er for, og 54 sier de er mot.

a) La *p* være sannsynligheten for at en tilfeldig valgt person støtter prosjektet. Angi et estimat for p, $(\hat{p})$ og tilhørende standardavvik (standardfeil), $SE(\hat{p})$.

b) Lag et 95 % konfidensintervall for p basert på dataene over.

Oppgave 4

Anta at antall timer per uke som høgskoleansatte trener er normalfordelt med forventing $\mu = 3{,}5$ og standardavvik $\sigma = 0{,}8$

a) Hva er sannsynligheten for at en tilfeldig valgt høgskoleansatt trener mer enn 4 timer per uke?

b) Hva er sannsynligheten for at 10 tilfeldig valgte høgskoleansatte i gjennomsnitt trener mindre enn 3 timer per uke ?

c) Hva er sannsynligheten for at en tilfeldig valgt høgskoleansatt trener mindre enn $2,5$ timer per uke ?

Du trekker ut 5 tilfeldige høgskoleansatte.

d) Hva er sannsynligheten for at akkurat 2 av disse 5 trener mindre enn 2,5 timer pr. uke?

e) Hva er sannsynligheten for at minst 1 av de 5 trener mindre enn 2,5 timer pr. uke?

Oppgave 5

Du har spurt $n = 20$ HiØ-utdannede økonomer om de får telefonkostnader dekket hos arbeidsgiver eller ikke. Undersøkelsen viste at 6 fikk dekket telefon.

a) Foreta en hypotesetest for å avgjøre om det er et mindretall (mindre enn halvparten) av Hiof-økonomer som får dekket telefonkostnadene av arbeidsgiver.

b) Beregn P-verdien (signifikanssannsynligheten) til resultatet av hypotesetesten i a).

Oppgave 6

Anta at du har undersøkt hvor mange timer pr. uke studenter er på sosiale medier, og at undersøkelsen din gav disse resultatene:

$\bar{X} = 6{,}00$ $\quad S_X = 3{,}0551$ $\quad n = 20$

a) Beregne et 95 % konfidensintervall for det gjennomsnittlige antall timer studenter er på sosiale medier pr. uke.

Anta at en undersøkelse viser at norske økonomistudenter i gjennomsnitt bruker 7 timer til jobbing med matematikkfag i uka. Legg til grunn at vi har gjort en tilsvarende undersøkelse ved HiØ, og fått følgende resultater:

$\bar{X} = 8{,}5$ $\quad S_X = 5{,}35$ $\quad n = 20$

b) Sett opp hypoteser og gjennomfør en test for å undersøke om studentene i Østfold bruker mer tid på matematikkfag enn landsgjennomsnittet. Brukt 5 % signifikansnivå.

Oppgave 7

Vi vurderer prisen på to varer mot hverandre. X er prisen på vare A og Y er prisen på vare B. Simultanfordelingen mellom de to er vist i denne tabellen:

		Y = Pris for B	
		34 kr	**24 kr**
X = Pris for A	**20 kr**	0,25	0,15
	30 kr	0,45	0,15

a) Finn marginalsannsynlighetene for X og Y.

b) Finn E[X] og E[Y].

c) Finn Var[X] og Var[Y].

d) Finn Cov[X,Y] og $\rho[X, Y]$.

Oppgave 8

Du skal plassere penger i 6 aksjefond av totalt 30 tilgjengelige. Du har ingen informasjon om fondene og vi antar at du velger helt tilfeldig.

a) Hvor mange mulige kombinasjoner av fond kan du sette sammen?

b) Etter et år gjør du opp status på investeringen og får vite hvilket fond som var best. Hva er sannsynligheten for at du hadde med det beste fondet i ditt tilfeldige utvalg?

c) Hva er sannsynligheten for at du hadde med deg de to beste fondene?

Vedlegg 1: Formelsamling

Grunnleggende formler i sannsynlighetsregningen

Komplementregel	$P(\bar{A}) = 1 - P(A)$
Generell addisjonssetning	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Betinget sannsynlighet	$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$
Multiplikasjonsregel	$P(A \cap B) = P(B \cap A) = P(B) \cdot P(A\|B) = P(A) \cdot P(B\|A)$
Bayes lov	$P(B \mid A) = \dfrac{P(B) \cdot P(A \mid B)}{P(A)}$
Total sannsynlighet	$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$
Uavhengighet	$P(A \cap B) = P(A) \cdot P(B)$ For to uavhengige begivenheter A og B gjelder: $P(A \mid B) = P(A)$ $P(B \mid A) = P(B)$

Kombinatorikk

La *n* være antall mulige utfall i én trekning, og *k* antall trekninger.

Ordnet utvalg med tilbakelegging	n^k
Ordnet utvalg uten tilbakelegging	$n_k = P_{n,k} = \dfrac{n!}{(n-k)!}$
Uordnet utvalg uten tilbakelegging	$\binom{n}{k} = C_{n,k} = \dfrac{n!}{(n-k)! \cdot k!}$

Generelt om sannsynlighetsfordelinger

Fordelingsfunksjon	$F(x) = P(X \leq x)$ $P(a < X \leq b) = F(b) - F(a)$ $P(X > a) = 1 - F(a)$ $P(X \leq b) = F(b)$
Forventning	$E(X) = \sum_{alle\ x_i} x_i \cdot P(X = x_i)$ $E(a) = a$ $E(bX) = bE(X)$ $E(a + bX) = a + bE(X)$ $E(a + bX + cX^2) = a + bE(X) + cE(X^2)$

	$E[g(X)] = \sum_{alle\ x_i} g(x_i) \cdot P(X = x_i)$
Varians	$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ $Var(X) = E[(X-\mu)^2] = E(X^2) - \left(E(X)\right)^2$ $Var(X+a) = Var(X)$ $Var(bX) = b^2 Var(X)$ $Var(bX+a) = b^2 Var(X)$
Standardavvik	$S_X = \sqrt{S_X^2}$ $\sigma[X] = \sqrt{Var(X)}$
Kovarians	$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y})$ $Cov(X,Y) = E[(X - E(X))(Y - E(Y)] = E(X \cdot Y) - E(X) \cdot E(Y)$
Korrelasjon	$R_{XY} = \frac{S_{XY}}{S_X \cdot S_Y}$ $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(Y)}\,\sqrt{Var(X)}}$

Diskrete sannsynlighetsfordelinger

Binomisk fordeling	$X \sim bin(n, p)$ $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ $E(X) = np \qquad Var(X) = np(1-p)$
Hypergeometrisk fordeling	$X \sim hypergeom(N, M, n)$ $P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$ $E(X) = n \cdot \frac{M}{N} \qquad Var(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)$
Poiossonfordeling	$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$ $E(X) = \lambda \qquad Var(X) = \lambda$

Kontinuerlige sannsynlighetsfordelinger

Generell normalfordeling	$X \sim N(\mu, \sigma^2)$ $F(x) = G\left(\frac{x-\mu}{\sigma}\right)$ $P(X \leq x) = P(Z \leq z) = G(z)$
Standard normalfordeling	$Z \sim N(0, 1)$ $Z = \frac{X-\mu}{\sigma}$ $\quad$ $P(Z \leq z) = G(z)$ $G(-z) = 1 - G(z)$

Tilnærminger

Sentralgrenseteoremet	La $X_1, X_2, \dots, X_n$ være uavhengige variabler fra samme fordeling med forventning µ og varians σ^2. Da er $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ $\quad$ tilnærmet $N\left(\mu, \frac{\sigma^2}{n}\right)$ og summen $X_1 + X_2 + \cdots + X_n$ tilnærmet $N(n\mu, n\sigma^2)$

Punktestimering

Estimering av µ	$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n} \qquad SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
Estimering av σ^2	$\widehat{\sigma^2} = S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad E(S_X^2) = \sigma^2$
Estimering av binomisk p	$\hat{p} = \frac{X}{n} \qquad SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Konfidensintervall

Z-intervall (kjent σ) for µ når n er stor (ca ≥ 30)/σ antas kjent	$\bar{X} \pm z_{\frac{\alpha}{2}} * \frac{\sigma_X}{\sqrt{n}}$
T-intervall for µ år n er liten (ca $< 30/S_X$ estimeres)	$\bar{X} \pm t_{\frac{\alpha}{2}}^{(v)} * \frac{S_X}{\sqrt{n}}$ $v = n - 1$
Konfidensintervall for p	$\left[\hat{p} \pm z_{\frac{\alpha}{2}} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \qquad \hat{p} = \frac{X}{n}$

Hypotesetesting

Z-test av μ når n er stor (ca $\geq$ 30)/σ antas kjent)	$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
T-test av μ år n er liten (ca $<$ 30/ S_X estimeres)	$T = \frac{\bar{X} - \mu_0}{\frac{S_X}{\sqrt{n}}}$
Z-test av p	$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Vedlegg 2: Tabeller

Tabell 4. Normalkurven (arealtabell)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.0000	–	–	–	–	–	–	–	–	–

Tabellen gir arealet $G(z)$ under normalkurven til venstre for z. Eksempel: $z = 1.54$ gir $G(z) = 0.9382$.

Tabell 5. Normalkurven (fraktiltabell)

a	z	a	z	a	z
0.50	0.0000	0,30	0.5244	0.10	1.2816
.49	0.0251	,29	0.5534	.09	1.3408
.48	0.0502	,28	0.5828	.08	1.4051
.47	0.0753	,27	0.6128	.07	1.4758
.46	0.1004	,26	0.6434	.06	1.5548
.45	0.1257	.25	0.6745	.050	1.6449
.44	0.1510	.24	0.7063	.045	1.6954
.43	0.1764	.23	0.7389	.040	1.7507
.42	0.2019	.22	0.7722	.035	1.8119
.41	0.2275	.21	0.8064	.030	1.8808
.40	0.2534	.20	0.8416	.025	1.9600
.39	0.2793	,19	0.8779	.020	2.0538
.38	0.3055	,18	0.9154	.015	2.1701
.37	0.3319	,17	0.9542	.010	2.3264
.36	0.3585	,16	0.9945	.005	2.5758
.35	0.3853	,15	1.0364	.0010	3.0902
.34	0.4125	,14	1.0803	.0005	3.2905
.33	0.4399	,13	1.1264	.0001	3.7190
.32	0.4677	,12	1.1750		
.31	0.4959	,11	1.2265		

Tabellen gir z slik at arealet til høyre for z under normalkurven er lik a, det vil si $a = P(X \geq z)$, der X er standard normalfordelt. Eksempel: $a = 0.045$ gir $z = 1.6954$.

Tabell 8. t-kurver (fraktiltabell)

ν \ a	0.25	0.10	0.05	0.025	0.010	0.005	0.001	0.0005
1	1.000	3.078	6.314	12.71	31.82	63.66	318.3	636.6
2	0.816	1.886	2.920	4.303	6.965	9.925	22.33	31.60
3	0.765	1.638	2.353	3.182	4.541	5.841	10.21	12.92
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
35	0.682	1.306	1.690	2.030	2.438	2.724	3.340	3.591
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
45	0.680	1.301	1.679	2.014	2.412	2.690	3.281	3.520
50	0.679	1.299	1.676	2.009	2.403	2.678	3.261	3.496
55	0.679	1.297	1.673	2.004	2.396	2.668	3.245	3.476
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460
70	0.678	1.294	1.667	1.994	2.381	2.648	3.211	3.435
80	0.678	1.292	1.664	1.990	2.374	2.639	3.195	3.416
90	0.677	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.677	1.290	1.660	1.984	2.364	2.626	3.174	3.390
∞	0.675	1.282	1.645	1.960	2.326	2.576	3.090	3.290

Tabellen gir t slik at arealet til høyre for t under t-kurven med ν frihetsgrader er lik a.
Eksempel: $\nu = 10$, $a = 0.005$, gir $t = 3.169$.