

UNIVERSITETET I OSLO ØKONOMISK INSTITUTT

Eksamen: **ECON2130 – Statistikk 1**

Eksamensdag: 29.05.2019

Sensur kunngjøres: 19.06.2019

Tid for eksamen: kl. 09:00 – 12:00

Oppgavesettet er på 5 sider

Tillatte hjelpemidler:

- Alle trykte og skrevne hjelpemidler er tillatt.
- To kalkulatorer av typen:
Aurora HC106 og/eller Casio FX-85EX

Eksamen blir vurdert etter ECTS-skalaen. A-F, der A er beste karakter og E er dårligste ståkarakter. F er ikke bestått.

Besvar alle oppgavene. Hver deloppgave har lik vekt.

Oppgave I

En kommune skal bygge ny idrettshall og vurderer to entreprenører, A og B. Begge gir samme pristilbud, men kommunen er bekymret for forsinkelser. For å undersøke hvem av de to som er best samler de inn informasjon om hvilke forsinkelser de to firmaene har hatt når de har bygget for andre kommuner i fylket. De finner følgende hyppigheter:

	A	B
Ferdig før tiden	4	7
Ferdig til avtalt tid	12	7
Forsinket 1 måned	2	6
Forsinket 2 måneder	4	4
Forsinket 3 måneder	6	2
Forsinket 4 måneder	0	2

I denne oppgaven skal vi anta at kolonne A gir den sanne fordelingen for entreprenør A, og likeså for kolonne B. Dermed ser vi bort fra eventuelle inferens-problemer.

- Hva er sannsynligheten for at entreprenør A er ferdig til avtalt tid eller før? Og hvor stor er denne sannsynligheten for entreprenør B?
- Gitt at et firma ble forsinket tre måneder eller mer, hva er sannsynligheten for at det var firma A når kommuner i utgangspunktet velger firma A og B med sannsynlighet $\frac{1}{2}$ hver?
- Definer to stokastiske variable X_A og X_B som antall måneder forsinkelse med bedrift A og B. Både ferdig før tiden og til avtalt tid regnes som 0.
Finn fordelingen til de to variablene og tegn et histogram for hver av dem.
- Finn forventning, varians og standardavvik til X_A og X_B .
- Beregn $E(X_A | X_A > 0)$ og forklar hva den betyr. Sammenlikn med tilsvarende for drift B.
- Ved forsinkelse kan kommunen ilegge dagbøter på kr 100 000 per måned. Definer stokastiske variable Y_A og Y_B som beløp betalt i dagbøter. Finn forventning og varians til Y_A og Y_B .
- Kommunen vurderer å bestille tre bygg (idrettshall, kulturhus og eldresenter) fra samme entreprenør. De ønsker å finne sannsynligheten for at total forsinkelse for de tre prosjektene blir over 6 måneder. Forklar hvordan du vil lage en simulering som beregner dette for hver av de to bedriftene, og forklar hvert enkelt skritt i algoritmen din.
Du kan svare med å sette opp et R-skript eller en algoritme i pseudo-kode. Du kan se på hvert av de tre byggeprosjektene som uavhengige av hverandre.

Oppgave II

I denne oppgaven skal vi se på lønnsinntekt. Det er en vanlig antakelse at logaritmen til inntekten er normalfordelt. Vi definerer en stokastisk variabel X som den naturlige logaritmen av månedsinntekten i dollar til en familie og antar $X \sim N(\mu, \sigma^2)$.

- Vi ønsker å estimere den ukjente μ ved hjelp av gjennomsnittet $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$. Dette er en forventningsrett estimator på μ . Forklar hva det betyr at estimatoren er forventningsrett og hvorfor det er en heldig egenskap.
- Anta at $\mu = 6.6$, $\sigma = 0.7$ og at vi har et utvalg på $n = 100$ familier. Hva er sannsynligheten for å observere $\bar{X} < 6.5$?
- En alternativ estimator for parameteren μ er medianen. Hvilke kriterier vil du legge til grunn for å velge hvilken av de to estimatorene gjennomsnitt og median som er best. Forklar hvordan du vil gjennomføre en simulering i R for å undersøke hvilken estimator som er best. Forklar hvert enkelt skritt i algoritmen din.

Du kan enten skrive et R-skript eller forklare algoritmen du vil bruke i pseudo-kode.

Videre i denne oppgaven skal vi bruke et utdrag fra den amerikanske NLSY-undersøkelsen¹. Vi har en data frame `dta` med data på log inntekt (`loginntekt`) og om noen i familien har lånekort på biblioteket (`bibkort`). Noen beregninger på dataene er gjengitt nedenfor:

```
> dstats = function(x) {
  c(mean=mean(x), sd=sd(x), n=length(x))
}

> dstats(dta$loginnt)
      mean      sd      n
6.5864336  0.7216815 920.0000000

> dstats(dta$loginnt[dta$bibkort==0])
      mean      sd      n
6.5302692  0.7182616 238.0000000
> dstats(dta$loginnt[dta$bibkort==1])
      mean      sd      n
6.606033   0.722368 682.0000000
```

- Vi skal til å begynne med se bort fra lånekort på biblioteket. Lag et estimat på parameteren μ og sett opp et 95 % konfidensintervall for estimatet. Vær klar på hvilke forutsetninger du gjør. Forklar hva konfidensintervallet betyr.
- Det er mye som tyder på at personer med bedre sosio-økonomisk bakgrunn gjør det bedre i arbeidsmarkedet enn de med dårligere bakgrunn. Det er blitt foreslått at en indikator på høy sosio-økonomisk status er at noen i familien har lånekort på biblioteket. Test hypotesen at forventet log inntekter er lik i gruppene med og uten lånekort. Du kan selv

¹ Se <https://www.bls.gov/nls/nlsy79.htm> for detaljer om undersøkelsen.

velge signifikansnivå. Vær klar på hvilke hypoteser du setter opp og hvilke forutsetninger du gjør.

- f) Beregn styrken på testen i e) hvis den sanne differansen er $\mu_{Med} - \mu_{Uten} = 0.1$. Du kan anta en felles kjent $\sigma = 0.72$. Hvordan tolker du svaret ditt?
- g) Gå tilbake til tilfellet uten å se på lånekort. Vi ønsker å finne forventet lønn ϕ , det vil si uten å ta logaritmer. Hvis vi fortsatt har X som log inntekten, er en foreslått estimator $\hat{\phi} = \exp(\bar{X})$ hvor \exp er eksponensialfunksjonen (dvs. den inverse av logaritmen) og \bar{X} gjennomsnittet av log inntekt. Bruk estimatoren til å estimere forventningen ϕ . Er denne estimatoren forventningsrett og/eller konsistent? Beregn også et 95 % konfidensintervall for ϕ .

TABLE 3: NORMAL CUMULATIVE DISTRIBUTION FUNCTION

For $Z \sim N(0,1)$, the function tabulated is $\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt$.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.02	0.9783
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.04	0.9793
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.06	0.9803
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.08	0.9812
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.10	0.9821
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.12	0.9830
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.14	0.9838
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.16	0.9846
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.18	0.9854
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	2.20	0.9861
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	2.22	0.9868
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	2.24	0.9875
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	2.26	0.9881
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	2.28	0.9887
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	2.30	0.9893
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	2.32	0.9898
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	2.34	0.9904
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	2.36	0.9909
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	2.38	0.9913
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	2.40	0.9918
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	2.42	0.9922
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	2.44	0.9927
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	2.46	0.9931
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	2.48	0.9934
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	2.50	0.9938
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	2.55	0.9946
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	2.60	0.9953
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	2.65	0.9960
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	2.70	0.9965
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	2.75	0.9970
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649	2.80	0.9974
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656	2.85	0.9978
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664	2.90	0.9981
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671	2.95	0.9984
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678	3.00	0.9987
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686	3.05	0.9989
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693	3.10	0.9990
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699	3.15	0.9992
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706	3.20	0.9993
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713	3.25	0.9994
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719	3.30	0.9995
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726	3.35	0.9996
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732	3.40	0.9997
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738	3.50	0.9998
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744	3.60	0.9998
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750	3.70	0.9999
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756	3.80	0.9999
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761	3.90	1.0000
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767	4.00	1.0000
0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772		

Royal Statistical Society Statistical Tables

TABLE 6: PERCENTAGE POINTS OF STUDENT'S t DISTRIBUTION

The values in the table are those which a random variable with Student's t distribution on ν degrees of freedom exceeds with the probability shown.

ν	0.100	0.050	0.025	0.010	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
45	1.301	1.679	2.014	2.412	2.690	3.281	3.520
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
55	1.297	1.673	2.004	2.396	2.668	3.245	3.476
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
70	1.294	1.667	1.994	2.381	2.648	3.211	3.435
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
90	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
110	1.289	1.659	1.982	2.361	2.621	3.166	3.381
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291