ECON2130: EKSAMEN 2013

TALLSVAR.
Det anbefales at de 9 deloppgavene merket med A, B,… teller likt – uansett variasjon i vanskelighetsgrad. Svarene er gitt i << … >>.

Oppgave 1

Et større amerikansk fagbok-forlag som blant annet publiserer lærebøker i statistikk, deler sine publiserte lærebøker i statistikk inn i 4 (disjunkte) suksess-kategorier ettersom hvordan bøkene lykkes i markedet. Suksess-kategoriene er: S_1 (meget suksessfull), S_2 (suksessfull), S_3 (middels suksessfull) og S_4 (mislykket). Basert på lengre tids markedserfaring fra tidligere publiserte lærebøker i statistikk opererer forlaget med følgende sannsynligheter (se tabell 1) for forskjellige suksess-kategorier for et vilkårlig nytt bokmanuskript som kommer inn til forlaget for publikasjon:

Tabell 1

Kategori	S_1	S_2	S_3	S_4
$P(S_j)$	0.10	0.20	0.40	0.30

Før publisering av et nytt manuskript lar alltid forlaget manuskriptet gjennomgå en fagfellevurdering. Resultatet av denne klassifiseres som enten "god" (G) eller "svak" (dvs "ikke-god", eller med symbol, $\bar{G}$). Basert på erfaring fra publiserte lærebøker anslår forlaget følgende sannsynligheter (se tabell 2) for G i de forskjellige suksess-kategoriene:

Tabell 2

Kategori	S_1	S_2	S_3	S_4
$P(G \mid S_j)$	0.99	0.70	0.40	0.20

A. **i.** Bestem $P(S_3 \cup S_4)$ og $P(S_3 \cap S_4)$.

ii. Forklar ved Venn-diagram, eller på annen måte, at

$$G \cap [S_3 \cup S_4] = (G \cap S_3) \cup (G \cap S_4)$$

iii. Finn $P(G \mid S_3 \cup S_4)$.

<< **Svar:** **i.** Siden S_3 og S_4 er disjunkte begivenheter, følger at $P(S_3 \cap S_4) = 0$, og $P(S_3 \cup S_4) = P(S_3) + P(S_4) = 0.7$.

ii. Venn-diagram …..

iii.

$$P(G \mid S_3 \cup S_4) = \frac{P(G \cap S_3) + P(G \cap S_4)}{P(S_3 \cup S_4)} = \frac{P(S_3)P(G \mid S_3) + P(S_4)P(G \mid S_4)}{P(S_3 \cup S_4)} =$$

$$= \frac{0.16 + 0.06}{0.7} = 0.314$$

>>

B. Anta at et nytt bok-manuskript som kommer inn til forlaget faktisk får god fagfellevurdering (G). Hva blir da sannsynlighetene for at boka faller i de forskjellige suksess-kategoriene i markedet? (**Hint:** Finn sannsynlighetene $P(S_j \mid G)$ for $j = 1, 2, 3, 4$.)

Svar: Multiplikasjonssetningen brukt på tabell 1 og 2 gir

Kategori	S_1	S_2	S_3	S_4	Sum $= P(G)$
$P(G \cap S_j)$	0.099	0.14	0.16	0.06	0.459

der, pga disjunkt union, $P(G) = \sum_{j=1}^{4} P(G \cap S_j)$ =0.459.

Dermed blir $P(S_j \mid G) = \dfrac{P(S_j \cap G)}{0.459}$, som gir

Kategori	S_1	S_2	S_3	S_4	Sum
$P(S_j \mid G)$	0.22	0.31	0.35	0.13	1.01 (pga avrunding)

>>

Oppgave 2

Regjeringen har tilsynelatende hatt en tendens til å anslå lønnsveksten i Norge lavere enn den faktisk realiserte lønnsveksten. Aftenposten hadde et oppslag om dette i fjor, 30. mai 2012, der det blant annet sto,

Sitat Aftenposten: *Regjeringen og Finansdepartementet har lang praksis i å spå for lav lønnsvekst. Finansminister Sigbjørn Johnsen har i sine reviderte nasjonalbudsjetter undervurdert lønnsveksten hvert år siden 2006.*

Artikkelen underbygger påstanden med data gitt i tabell 3, der variabelen x betegner regjeringens spådom et vilkårlig år, y den realiserte lønnsveksten samme år og d differansen $y - x$.

Tenk deg at du befinner deg i mai 2012 og at den faktiske lønnsveksten for 2012 ennå ikke er kjent.

Tabell 3 Lønnsvekst[1] i %

Observasjon i	År	Slik spådde regjeringen x_i	Slik gikk det y_i	Differanse $d_i = y_i - x_i$
1	2000	3.75	4.4	0.65
2	2001	4.50	4.8	0.30
3	2002	5.00	5.7	0.70
4	2003	4.50	4.5	0.00
5	2004	3.75	3.5	-0.25
6	2005	3.25	3.3	0.05
7	2006	3.50	4.1	0.60
8	2007	4.75	5.4	0.65
9	2008	5.50	6.3	0.80
10	2009	4.00	4.2	0.20
11	2010	3.25	3.7	0.45
12	2011	3.90	4.2	0.30
13	2012	3.75	-	-

For å lette regningen senere i oppgaven oppgis noen deskriptive størrelser i tabell 4.

[1] **Kilde.** Revidert nasjonalbudsjett 2000-2012 og rapporter fra *Det Tekniske beregningsutvalg for inntektsoppgjørene* 2012 og 2010.

Tabell 4 Noen deskriptive størrelser for perioden 2000 – 2011

n (Antall observasjoner)	12
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	4.1375
$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$	4.5083
$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$	0.3708
$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$	0.5087
$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y})^2$	0.8190
$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i-\overline{d})^2$	0.1066
$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$	0.6106

A. **i.** Beregn den empiriske korrelasjonskoeffisienten, *r*, mellom *x* og *y* basert på data i tabell 3.
ii. Gjør kort rede for hva *r* uttrykker.

<< Svar:

i. $r = \frac{s_{xy}}{s_x s_y} = \frac{0.6106}{\sqrt{(0.5087)\cdot(0.8190)}} = 0.9459$

ii. …. >>

B. Anta nå at x_i, y_i og d_i er observasjoner av stokastiske variable, X_i, Y_i og $D_i = Y_i - X_i$, der *parene* $(X_i, Y_i), \quad i = 1, 2, \ldots, n$ antas å være uavhengige av hverandre[2] og identisk fordelte. Betegn forventningene med μ_X, μ_Y og μ_D henholdsvis og variansene med σ_X^2, σ_Y^2 og σ_D^2 henholdsvis. I tillegg antar vi at $D_1, D_2, \ldots, D_n$ er uavhengige og identisk normalfordelte, $D_i \sim N(\mu_D, \sigma_D)$ for $i = 1, 2, \ldots, n$.

[2] Dette innebærer at X_i -er og Y_i -er fra forskjellige par er uavhengige. X_i og Y_i fra samme par kan imidlertid være avhengige.

i. Forklar hvordan sammenhengen $\bar{d} = \bar{y} - \bar{x}$ i tabell 4 følger av regneregler for summer brukt på tall i tabell3.
ii. Forklar hvorfor $\mu_D = \mu_Y - \mu_X$.
iii. Synes du at det ville være rimelig å forutsette en sammenheng $\sigma_D^2 = \sigma_Y^2 + \sigma_X^2$ mellom variansene, eller synes du ikke det? Begrunn svaret ditt.

<< **Svar: i.** Av regnereglene for summer i appendikset i regresjon-I-notatet følger

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) = \frac{1}{n}\left[\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{y} - \bar{x}$$

ii. $\mu_D = E(D_i) = E(Y_i - X_i) \overset{\text{regel 4.12}}{=} E(Y_i) - E(X_i)$

iii. $\sigma_D^2 = \sigma_Y^2 + \sigma_X^2$ forutsetter at X_i og Y_i er ukorrelerte i populasjonen. I den foreliggende modellen kan *r* i punkt A oppfattes som et estimat for korrelasjonen mellom X_i og Y_i, som er så stor at forutsetningen virker høyst urimelig.
>>

C. **i.** Sett opp en forventningsrett estimator for μ_D og begrunn forventningsrettheten.
ii. Sett opp en test for hypotesen $H_0 : \mu_D \le 0$ mot $H_1 : \mu_D > 0$ med signifikansnivå 5%.
iii. Gjennomfør testen ut fra data i tabell 3 og 4 og formuler en konklusjon. Kommenter resultatet.

<< **Svar: i.** $\hat{\mu}_D = \bar{D}$ er forventningsrett siden regel 4.12 gir

$$E(\bar{D}) = \frac{1}{n}E\left(\sum D_i\right) = \frac{1}{n}\sum E(D_i) = \frac{1}{n}n\mu_D = \mu_D$$

ii. **Testobservator:** $T = \dfrac{\hat{\mu}_D}{SE(\hat{\mu}_D)} = \dfrac{\bar{D}}{S_D/\sqrt{n}}$ er *t*-fordelt med $n-1=11$ frihetsgrader hvis $\mu_D = 0$. 5%-kvantilen i t_{11}-fordelingen er 1.796.

Test: Forkast H_0 hvis $T > 1.796$

iii. $\hat{\mu}_{D,obs} = \bar{D}_{obs} = 0.3708$, $SE(\hat{\mu}_D) = S_D/\sqrt{n} = 0.0942$, som gir $T_{obs} = 3.934$.
Konklusjon: Forkast H_0. **Kommentar:** Resultatet bekrefter påstanden i artikkelen.

>>

D. Vi ønsker å utnytte denne tendensen til å korrigere regjeringens spådom (3.75) for 2012. En enkel måte å gjøre dette er å postulere en regresjonsmodell for Y_i med x_i som (ikke-stokastisk) forklaringsvariable. Vi antar altså

(1) $Y_i = \alpha + \beta x_i + e_i$ for $i = 1, 2, \ldots, n$

der restleddene $e_1, e_2, \ldots, e_n$ antas uavhengige og identisk normalfordelte med forventning 0 og varians σ^2.

i. La *Y* betegne realisert lønnsvekst et vilkårlig år der regjeringen spår en lønnsvekst på *x*%. Beregn et estimat for regresjonsfunksjonen, $E(Y) = \mu(x) = \alpha + \beta x$, basert på minste kvadraters estimatorer for α og β.
ii. Bruk regresjonsanalysen i punkt **i.** til å foreslå et korrigert estimat på forventet lønnsvekst i 2012, basert på regjeringens spådom 3.75.

<< **Svar: i.**

$$\hat{\beta} = \frac{S_{xY}}{s_x^2} \to \frac{0.6106}{0.5087} = 1.2003, \qquad \hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x} \to 4.5083 - (1.2)(4.1375) = -0.4567$$

Den estimerte regresjonsfunksjonen blir $\hat{\mu}(x) = -0.4567 + (1.2) \cdot x$

ii. Korrigert estimat på forventet lønnsvekst i 2012:
$\hat{\mu}(3.75) = -0.4567 + (1.2)(3.75) = 4.04$

>>

E. Beregn et 95% konfidensintervall for forventet lønnsvekst i 2012 basert på regresjonsmodellen og ditt korrigerte estimat i punkt **D ii**.[3]

<< **Svar:**

Siden $T = \dfrac{\hat{\mu}(3.75) - \mu(3.75)}{SE(\hat{\mu}(3.75))} \sim t_{10}$-fordelt, blir konfidensintervallet

$$\hat{\mu}(3.75) \pm t_{10,\,0.025} SE\left(\hat{\mu}(3.75)\right) = \hat{\mu}(3.75) \pm (2.228) \cdot SE\left(\hat{\mu}(3.75)\right)$$

---

[3] Som kuriositet kan nevnes at *Det tekniske beregningsutvalget for inntektsoppgjørene* senere på året fastsatte gjennomsnittlig lønnsvekst i 2012 til 4%.

der $$SE\left(\hat{\mu}(3.75)\right) = \begin{cases} \hat{\mu}(3.75) \pm (2.228)\cdot\hat{\sigma}\cdot\sqrt{\frac{1}{n}+\left(\frac{3.75-\bar{x}}{\hat{\sigma}/SE(\hat{\beta})}\right)^2} & \text{Løvås regel 7.5} \\ \hat{\mu}(3.75) \pm (2.228)\cdot\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(3.75-\bar{x})^2}{(n-1)s_x^2}} & \text{HG regresjonsnotat II} \end{cases}$$

og

$$\hat{\sigma}^2(= s^2 \text{ i Løvås}) = \frac{SSE}{n-2} \overset{\text{regresjonsnotat II}}{=} \frac{n-1}{n-2}\left(S_y^2 - \hat{\beta}^2 s_x^2\right) \to \frac{11}{10}\left(0.819-(1.2)^2(0.5087)\right) = 0.09512$$

$$\hat{\sigma} \to 0.3084$$

Dermed $SE\left(\hat{\mu}(3.75)\right) = \hat{\sigma}\sqrt{\frac{1}{n}+\frac{(3.75-\bar{x})^2}{(n-1)s_x^2}} = (0.3084)\sqrt{\frac{1}{12}+\frac{(3.75-4.1375)^2}{11\cdot 0.5087}} = 0.1024$

som gir 95% KI for $\mu(3.75)$:

$$\hat{\mu}(3.75) \pm (2.228)\cdot SE\left(\hat{\mu}(3.75)\right) \to 4.04 \pm 0.22807 = [3.81,\ 4.27]$$

\>>

Oppgave 3

En annen måte å se på regjeringens tendens til å undervurdere lønnsveksten, er å betrakte hvert år som et binomisk forsøk der vi definerer suksessbegivenheten (*S*) som begivenheten at regjeringens spådom ligger under den realiserte lønnsveksten. Anta at sannsynligheten for at regjeringens spådom ligger under den realiserte, er konstant lik *p*, og at *S* i forskjellige år er uavhengig av hverandre. La *X* være antall ganger *S* inntreffer i løpet av en periode på *n* år. Vi antar således at *X* er binomisk fordelt, $X \sim \text{bin}(n, p)$. I denne modellen kan tendensen til å undervurdere lønnsveksten uttrykkes ved hypotesen $p > 0.5$, mens fravær av en slik tendens uttrykkes ved $p = 0.5$

I denne oppgaven skal vi anta $n = 10$ år (dvs. kun se på perioden 2002-2011). (Dette for å kunne utnytte tabellene i Løvås.)

A. Vi ønsker å teste $H_0 : p \le 0.5$ mot $H_1 : p > 0.5$ basert på *X*. Følgende test blir foreslått:

(2) Forkast H_0 hvis $X > 7$.

i. Beregn signifikansnivået for testen i (2) tilnærmet basert på normaltilnærmelsen til binomiske fordelinger med heltallskorreksjon.
ii. Beregn signifikansnivået for testen i (2) eksakt ved bruk av tabeller over den binomiske fordelingen.
iii. Gjennomfør testen basert på perioden 2002-2011 i tabell 3 (i oppgave 2) og formuler en konklusjon.

<< **Svar:** **i.** Nivået er gitt ved $\alpha = P_{p=0.5}(\text{forkast } H_0) = P_{p=0.5}(X > 7)$. Ved normaltilnærmelsen får vi

$$\alpha = P_{p=0.5}(X > 7) = 1 - P_{p=0.5}(X \le 7) \overset{heltallskorreksjon}{=} 1 - P_{p=0.5}(X \le 7.5)$$
$$\approx 1 - P\left(Z \le \frac{7.5 - 10(0.5)}{\sqrt{10(0.5)(0.5)}} \right) = 1 - G(1.58) = 1 - 0.9429 = 0.057$$

ii. Eksakt beregning ved tabell D.1 i Løvås:

$$\alpha = P_{p=0.5}(X > 7) = 1 - P_{p=0.5}(X \le 7) \overset{\text{tabell D.1}}{=} 1 - 0.945 = 0.055$$

iii. I følge tabell 3 er $X_{obs} = 9$ slik at konklusjonen blir: Forkast H_0. (Også $X_{obs} = 8$ bør aksepteres siden det er en observasjon d = 0, og *S* er ikke helt presis oppgaveteksten.)
\>>

B. Beregn både sannsynligheten for feil av type I og sannsynligheten for feil av type II hvis den sanne verdien av *p* er 0.4. Gjør det samme hvis den sanne *p* = 0.5 og hvis *p* = 0.7. Du kan selv velge om du vil bruke eksakt beregning (tabell D.1 i Løvås) eller normaltilnærmelsen.

<< **Svar:**

	Eksakt (tabell D.1)	Med normaltilnærming

p	$P(\text{forkast } H_0) = P_p(X > 7)$	P(feil I)	P(feil II)	z	$P(\text{forkast } H_0) \approx 1 - G(z)$
0.4	0.012	0.012	0	2.26	0.012
0.5	0.055	0.055	0	1.58	0.057
0.7	0.383	0	0.617	0.35	0.365

>>