UNIVERSITETET I OSLO
ØKONOMISK INSTITUTT

Eksamen i: **ECON2130 – Statistikk 1**
Exam: ECON2130 – Statistics 1

Eksamensdag: Fredag 23. mai 2008
Date of exam: Friday, May 23, 2008

Sensur kunngjøres: Torsdag 12. juni
Grades will be given: Thursday June 12

Tid for eksamen: kl. 09:00 – 12:00
Time for exam: 09:00 a.m. – 12:00 noon

Oppgavesettet er på X sider
The problem set covers X pages

English version on page

Tillatte hjelpemidler:
- Alle trykte og skrevne hjelpemidler, samt lommekalkulator er tillatt

Resources allowed:
- *All written and printed resources, as well as calculator is allowed*

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

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

Oppgave 1

Vi har en kortstokk bestående av 6 kort. På 2 av disse står det skrevet JA på forsiden mens det står NEI på forsiden av de 4 andre kortene. Hvis man får se kortet med baksiden vendt mot seg, er det ikke mulig å se om det står JA eller NEI på forsiden.

A. Anta kortstokken blir stokket godt og lagt ned på bordet i en bunke med baksiden opp.

(i) Hva er sannsynligheten for at det øverste kortet i bunken er et JA-kort?

(ii) Hva er sannsynligheten for at de to øverste kortene i bunken begge er JA-kort?

B. La generelt A og B være to begivenheter. Da gjelder

(1) $$A = (A \cap B) \cup (A \cap \overline{B})$$

(i) Forklar relasjonen (1) ved et Venn-diagram.

(ii) Anta $P(A) = \dfrac{4}{5}$ og $P(B \mid A) = \dfrac{1}{4}$. Hva blir da $P(A \cap \overline{B})$?

(iii) Anta at sannsynlighetene oppgitt i (ii) fortsatt gjelder. Hvilken verdi må $P(B)$ ha hvis A og B er uavhengige begivenheter? Hva blir i så fall $P(A \cup B)$?

C. Vi vender tilbake til situasjonen i punkt **A**. La J_1 være begivenheten at det øverste kortet i bunken er et JA-kort, og J_2 begivenheten at det nest øverste kortet er et JA-kort. Forklar intuitivt, eller ved en beregning, at $P(J_2) = P(J_1) = \frac{1}{3}$.

D. Kortstokken stokkes grundig og legges som før på bordet i en bunke med baksiden opp. Ett og ett kort åpnes deretter fra toppen av bunken og nedover inntil første JA-kort dukker opp. La X være antall kort som må snus inntil første JA-kort dukker opp. Hvis kortstokken er grundig nok stokket, vil alle mulige rekkefølger av kortene i bunken ha samme sannsynlighet, og den stokastiske variabelen X vil ha en fordeling gitt i tabell 1:

Tabell 1 *Fordelingen for X*

X	1	2	3	4	5	6
$P(X = x)$	$\frac{5}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	0

(i) Vis at $P(X = 3) = \frac{3}{15}$ som gitt i tabellen.

[**Hint:** Merk at begivenheten $(X = 3)$ er ekvivalent med begivenheten $N_1 \cap N_2 \cap J_3$ - dvs. der de to øverste kortene er NEI-kort og det tredje kortet fra toppen et JA-kort (der begivenheten N_i betyr at *i*-te kort fra toppen er et NEI-kort).]

(ii) Vis at $E(X) = \frac{7}{3}$ og $\mathrm{Var}(X) = \frac{14}{9}$.

E. Å stokke en kortstokk godt er vanskeligere enn folk flest tror. Jens mener at han vanligvis stokker kortene godt nok, men er villig til å gjennomføre følgende test: Eksperimentet beskrevet i punkt **D** gjentas $n = 40$ ganger. I hvert forsøk stokker Jens kortene så godt han mener er tilstrekkelig, legger deretter kortene på bordet i en bunke med baksiden opp, åpner ett og ett kort fra toppen og nedover i bunken inntil første JA-kort dukker opp og registrerer til slutt hvor mange kort som må snus før JA-kortet viser seg. La for forsøk i X_i betegne antall kort som må snus inntil det første JA-kortet kommer $(i = 1, 2, \ldots, 40)$.

Anta at $X_1, X_2, \ldots, X_{40}$ er uavhengige og identisk fordelt, med felles forventning, $E(X_i) = \mu$ og varians, $\mathrm{Var}(X_i) = \sigma^2$. Hvis stokkingen er god nok, er den felles fordelingen kjent og gitt i tabell 1. Dette utgjør vår null-hypotese. Hvis stokkingen ikke er god nok, vil den felles fordelingen være ukjent med ukjent forventning og varians.

Gjennomsnittlig *X*-verdi for de 40 stokke-forsøkene til Jens ble 1,98. Sett opp og gjennomfør (basert på dette resultatet) en test med signifikansnivå 5% for hypotesen:

$$H_0: \quad \mu = \frac{7}{3} = 2{,}33 \quad \text{mot} \quad H_1: \quad \mu \neq 2{,}33$$

Beregn p-verdien for testen og formuler en konklusjon.

> [**Hint:** Bruk sentralgrenseteoremet. Merk at både $\mu = \frac{7}{3}$ og $\sigma^2 = \frac{14}{9}$ er kjente under H_0, som i dette tilfellet er det eneste vi trenger å vite for å kunne beregne kritiske verdier og p-verdi.]

Oppgave 2

Et problem ved spørreundersøkelser er å få pålitelige svar på sensitive spørsmål. For å sikre anonymiteten benytter en forsker følgende intervju-metode. Anta spørsmålet er, "Har du kjøpt smuglersprit i år?". Forskeren viser respondenten (dvs. personen som blir intervjuet) en kortstokk av samme type som i oppgave 1 der JA står på 2 av kortene og NEI på 4. Respondenten får opplyst hvor mange av kortene som har JA og hvor mange som har NEI. Respondenten blir bedt om ikke å svare direkte på spørsmålet, men i stedet å trekke et kort og si om det som står på kortet stemmer eller ikke. For eksempel hvis respondenten faktisk ikke har kjøpt smuglersprit og trekker et kort med NEI på, så svarer vedkommende at det som står på kortet stemmer. Deretter blir kortet lagt tilbake i kortstokken uten at forskeren får se hva som står på det.

For enkelthets skyld vil vi i denne oppgaven se bort ifra de tilfellene der respondenten nekter å svare eller lyver, og antar i stedet at alle gir et av to mulige svar, "det stemmer", eller "det stemmer ikke", og at alle snakker sant.

A. Metoden brukes på et rent tilfeldig utvalg av voksne personer fra befolkningen. Gjør rede for at antall personer i utvalget som svarer, "det som står på kortet stemmer", kan antas å være binomisk fordelt.

B. La q være den relative andelen som ville ha svart, "det som står på kortet stemmer", dersom forskeren hadde intervjuet hele den voksne befolkningen. La p være andelen i befolkningen som har kjøpt smuglersprit. Det kan vises (jfr. punkt **D**) at sammenhengen mellom p og q er gitt ved

(2) $$q = \frac{2-p}{3}$$

La X være antall som svarer "det som står på kortet stemmer" i et tilfeldig utvalg på n voksne. Forklar hvorfor

$$\hat{p} = 2 - 3\frac{X}{n}$$

er en forventningsrett estimator for p. Vis at standardavviket (SD) for $\hat{p}$ er (uttrykt ved q):

$$\mathrm{SD}(\hat{p}) = 3\sqrt{\frac{q(1-q)}{n}}$$

C. (i) Konstruer et 95% konfidensintervall for p basert på $\hat{q} = \frac{X}{n}$.

[**Hint:** Ta utgangspunkt i et 95% konfidensintervall for q, som vi kan skrive (A, B), der A og B er stokastiske variable som oppfyller $P(A \le q \le B) = 0{,}95$ (tilnærmet). Dette intervallet kan du så overføre til et intervall for p ved å vise at begivenheten $(A \le q \le B)$ er ekvivalent med begivenheten $(2 - 3B \le p \le 2 - 3A)$.]

(ii) Hvor mange observasjoner trengs for at lengden, $L = B - A$, på konfidensintervallet for q ikke skal overskride 0,06? Beregn også hvor mange observasjoner som trengs for at lengden på intervallet for p ikke skal overskride 0,06.

D. Vis relasjonen (2) i punkt **B**.

[**Hint:** Betrakt intervju-situasjonen med en tilfeldig trukket respondent og innfør begivenhetene:

K = "Respondenten har kjøpt smuglersprit i år"
S = "Respondenten svarer at det som står på kortet stemmer"
J = " Det står JA på kortet som respondenten trekker"
N = " Det står NEI på kortet som respondenten trekker"

Gjør rede for og utnytt at $S = (K \cap J) \cup (\bar{K} \cap N)$, som er en disjunkt union. Merk også at det som står på kortet, åpenbart er uavhengig av om respondenten har kjøpt smuglersprit eller ikke.]

ENGLISH VERSION

Problem 1

We have a deck of cards consisting of 6 cards. On 2 of the cards the word YES is written on the front side while NO is written on the remaining 4 cards. A viewer looking at backside of a card cannot see whether YES or NO is written on the front side.

A. Suppose that the deck is shuffled well and put on the table in one pile with the backside up.

(i) What is the probability that the top card in the deck is a YES card?

(ii) What is the probability that the two top cards in the deck both are YES cards?

B. Let, in general, *A* and *B* be two events. Then it is always true that

(1) $$A = (A \cap B) \cup (A \cap \overline{B})$$

(i) Explain the relation (1) by a Venn diagram.

(ii) Suppose $P(A) = \frac{4}{5}$ and $P(B \mid A) = \frac{1}{4}$. What is then $P(A \cap \overline{B})$?

(iii) Suppose that the probabilities given in (ii) still hold. What must be the value of $P(B)$ if *A* and *B* are independent events? If so, what is $P(A \cup B)$?

C. We return to the situation in section **A**. Let J_1 be the event that the top card in the deck is a YES card, and J_2 the event that the second card from the top is a YES card. Explain intuitively, or by means of a calculation, that $P(J_2) = P(J_1) = \frac{1}{3}$.

D. The deck is shuffled well and put on the table as before, in a pile with the backside up. The cards in the deck are then turned one by one from the top until the first YES card appears. Let *X* be the number of cards that have to be turned until the first YES card shows up. If the deck of cards has been shuffled sufficiently well, all possible orderings of the 6 cards in the deck will have the same probability, and the random variable *X* will have a distribution as given in table 1:

Table 1 *The distribution of* X

X	1	2	3	4	5	6
$P(X = x)$	$\frac{5}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	0

(i) Show that $P(X = 3) = \frac{3}{15}$, as given in the table.

[**Hint:** Note that the event $(X = 3)$ is equivalent with the event $N_1 \cap N_2 \cap J_3$ - i.e., where the two top cards are NO cards and the third card from the top is a YES card. (The event N_i means that the *i*-th card from the top is a NO card.)]

(ii) Show that $E(X) = \frac{7}{3}$ and $\mathrm{Var}(X) = \frac{14}{9}$.

E. To shuffle a deck of cards well is more difficult than most people think. Jens thinks that he usually shuffles the cards well enough, but agrees to perform the following test: The experiment described in section *D* is repeated *n* = 40 times. In each trial Jens shuffles the cards as well as he thinks is sufficient, then puts the cards on the table in a pile with the backside up, turns the cards one by one from the top of the deck and downwards until the first YES card shows up, and writes down the number of cards that had to be turned until the first YES card appears. Let, for trial *i*, X_i denote the number of cards that has to be turned until the first YES card shows up $(i = 1, 2, \ldots, 40)$.

Suppose that $X_1, X_2, \ldots, X_{40}$ are independent and identically distributed, with common expectation, $E(X_i) = \mu$, and variance, $\mathrm{Var}(X_i) = \sigma^2$. If the shuffling is sufficiently good, the common distribution is known and given in table 1. This constitutes our null-hypothesis. If the shuffling is not good enough, the common distribution is unknown with unknown expectation and variance.

The mean *X*-value for the 40 shuffling-trials of Jens turned out to be 1,98. Set up and perform (based on this result) a test with level of significance 5% for the hypothesis:

$$H_0: \quad \mu = \frac{7}{3} = 2{,}33 \quad \text{versus} \quad H_1: \quad \mu \neq 2{,}33$$

Calculate the p-value for the test and formulate a conclusion.

> [**Hint:** Use the central limit theorem. Note that both $\mu = \frac{7}{3}$ and $\sigma^2 = \frac{14}{9}$ are known under H_0, which, in this case, is the only thing we need to know in order to calculate the critical values and p-value.]

Problem 2

A common problem with survey investigations is to obtain reliable answers to sensitive questions. To ensure the anonymity a scientist uses the following interview method. Assume that the question is, "Have you bought any bootleg (smuggled liquor) this year?". The scientist shows a deck of cards to the respondent (i.e., the person being interviewed) of the same type as in problem 1 where YES is written on 2 of the cards and NO on 4. The respondent is told how many cards have YES and how many have NO. The respondent is asked not to answer the question directly but to draw a card in stead and tell if what is written on the card is correct or not. For example, if the respondent actually did not buy any bootleg that year and draws a card with NO, then she or he answers that the card is correct. Then the card is put back into the deck in such a way that the scientist cannot see what is written on it.

For simplicity we will, in this problem, ignore those cases where the respondent refuses to answer or answers with a lie, and assume instead that everyone answer one of the two possible answers, ”the card is correct” or ”the card is not correct”, and that nobody lies.

A. The method is applied on a random sample drawn from the grown-up population. Explain why the number of people in the sample answering, "the card is correct", can be assumed to be binomially distributed.

B. Let q denote the relative frequency of people answering, "the card is correct", if the scientist had interviewed the complete grown-up population. Let p be the relative frequency of people in the population that have actually bought bootleg. It can be shown (see section **D**) that the relationship between p and q is given by

(2) $$q = \frac{2-p}{3}$$

Let X 1be the number of people answering "the card is correct" in a random sample of n grown-ups. Explain why

$$\hat{p} = 2 - 3\frac{X}{n}$$

is an unbiased estimator of p. Show that the standard deviation (SD) of $\hat{p}$ is (expressed by q):

$$\mathrm{SD}(\hat{p}) = 3\sqrt{\frac{q(1-q)}{n}}$$

C. (i) Construct a 95% confidence interval for p based on $\hat{q} = \frac{X}{n}$.

[Hint: Start with a 95% confidence interval for q, which we may write (A, B), where A and B are random variables that satisfy $P(A \le q \le B) = 0{,}95$ (approximately). You can then transfer this interval to an interval for p by showing that the event $(A \le q \le B)$ is equivalent to the event $(2-3B \le p \le 2-3A)$.]

(ii) How many observations are needed in order that the length, $L = B - A$, of the confidence interval does not exceed 0,06? Also calculate the number of observations needed in order that the length of the interval for p does not exceed 0,06.

D. Show the relation (2) in section **B**.

[**Hint:** Consider the interview situation with a randomly drawn respondent, and define the following events:

K = "The respondent has bought bootleg this year"
S = "The respondent answers that what is written on the card is correct"
J = "YES is written on the card that the respondent draws from the deck"
N = "NO is written on the card that the respondent draws from the deck"

Explain and utilize that $S = (K \cap J) \cup (\overline{K} \cap N)$, which is a disjoint union. Notice also that what is written on the card, evidently must be independent of the respondent having bought bootleg or not.]