UNIVERSITETET I OSLO
ØKONOMISK INSTITUTT

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

Eksamensdag: Mandag 23. mai 2005
Date of exam: Monday, May 23, 2005

Sensur kunngjøres: Tirsdag 14. juni
Grades will be given: Tuesday, June 14

Tid for eksamen: kl. 09:00 – 12:00
Time for exam: 9:00 a.m. – 12:00 noon

Oppgavesettet er på 7 sider
The problem set covers 7 pages

English version on page 5

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

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

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

I denne oppgaven kan du anta at sannsynligheten for at det blir pike (og sannsynligheten for at det blir gutt) i en vilkårlig fødsel er 1/2, og at kjønnet i forskjellige fødsler er uavhengige av hverandre. Vi er interessert i kjønnsfordelingen av barna for en vilkårlig tre-barns familie. Definer utfallet *xyz* som kjønnet til de tre barna der *x* er kjønnet til det eldste, *y* til det nest eldste og *z* til det yngste barnet. For eksempel, utfallet *PGP* betyr at det eldste og det yngste barnet er piker mens nest eldste er gutt.

a. Skriv opp utfallrommet som i dette tilfellet (for en tre-barns familie) består av 8 mulige utfall. Forklar hvorfor det er rimelig å anta at den tilhørende sannsynlighetsmodellen er uniform.

b. Definer begivenhetene (hendelsene):

A = “Familien har 2 piker og en gutt”
B = “Familien har minst 2 piker”
C = “Det eldste barnet er pike”

i) Finn P(A), P(B) og P(C).
ii) Finn $P(B \cup C)$, $P(B \cap C)$ og $P(B \mid C)$.

c. **i)** Om en tre-barns familie vet vi at de har minst to piker. Hva er i så fall sannsynligheten for at det tredje barnet er gutt? Merk at vi her ikke vet noe om det tredje barnet er født som nr. 1, 2 eller 3.
ii) Om en annen tre-barns familie vet vi at de to eldste barna er piker. Hva er da sannsynligheten for at det tredje barnet er gutt?

d. Et ungt par ønsker seg minst ett barn av hvert kjønn og bestemmer seg til å fortsette å få barn til målet er nådd. For enkelthets skyld antar vi at det første barnet de fikk var pike. La X betegne antall nye forsøk (barnefødsler) som må til inntil første gutt blir født. Etter at målet (minst ett av hvert kjønn) er nådd vil de ikke ha flere barn. Forklar hvorfor det er rimelig å anta at X er geometrisk fordelt når vi ser bort fra komplikasjoner som for eksempel ufruktbarhet, tvillingfødsler, og det faktum at det er begrenset hvor mange barn et par i praksis kan få.

La Y være antall barn paret ender opp med i alt. Med andre ord, $Y = X + 1$. Beregn forventning og varians for Y samt sannsynligheten for at Y blir minst lik 7.

e. Etter å ha tenkt seg om finner paret ut at de uansett ikke vil ha mer enn maksimum 6 barn. Med andre ord, selv om de ender opp med 6 piker, så vil de slutte med å få flere barn. Sett opp sannsynlighetsfordelingen til Y (antall barn i alt, inklusive den førstefødte piken) i dette tilfellet. Det vil si, bestem de aktuelle sannsynlighetene i følgende tabell:

y	2	3	4	5	6
$P(Y = y)$	?	?	?	?	?

Hva blir forventet Y i dette tilfellet?

Oppgave 2

Normaltemperaturen hos mennesker varierer litt fra person til person. Inntil relativt nylig har man regnet gjennomsnittlig normaltemperatur i populasjonen av mennesker som $37,0$ grader Celcius. Tabell 1 viser resultatet av måling av normaltemperaturen for et utvalg av $n = 130$ personer, ordnet i stigende rekkefølge. (For hvert individ er normaltemperaturen bestemt som et gjennomsnitt av flere målinger over tid.) Merk at tabellen består av 10 rader og 13 kolonner.

Tabell 1 (Normaltemperatur for 130 personer)

35.7	35.8	35.9	35,9	36,0	36,1	36,1	36,2	36,2	36,2	36,2	36,2	36,2
36,3	36,3	36,3	36,3	36,3	36,3	36,4	36,4	36,4	36,4	36,4	36,4	36,5
36,5	36,5	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6
36,6	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7
36,7	36,7	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8
36,8	36,8	36,8	36,8	36,9	36,9	36,9	36,9	36,9	36,9	36,9	36,9	36,9
36,9	36,9	36,9	37,0	37,0	37,0	37,0	37,0	37,0	37,0	37,0	37,0	37,0
37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1
37,1	37,1	37,1	37,1	37,1	37,2	37,2	37,2	37,2	37,2	37,2	37,2	37,3
37,3	37,3	37,3	37,3	37,3	37,4	37,4	37,4	37,4	37,5	37,7	37,8	38,2

Gjennomsnitt: 36,81 Standardavvik: 0,4074

a. Bestem medianen for observasjonene. Spiller det noen rolle om gjennomsnitt og standardavvik er beregnet på grunnlag av de sorterte observasjonene eller på grunnlag av de opprinnelige (usorterte) dataene? Gi en grunn for ditt svar.

b. Observasjonene er gruppert i intervaller i tabell 2. Merk at øvre grense i hvert intervall er inklusiv, mens nedre grense er eksklusiv. Dette betyr, for eksempel, at observasjonen 36,1 hører med i intervall nr. 2 og ikke i intervall nr. 3.

Tabell 2

Intervall	Frekvens
35,5 - 35,8	2
35,8 - 36,1	5
36,1 - 36,4	18
36,4 - 36,7	29
36,7 - 37,0	37
37,0 - 37,3	31
37,3 - 37,6	?
37,6 - 37,9	?
37,9 - 38,2	?
Sum	130

Fyll inn de korrekte frekvensene der det står spørsmålstegn i tabellen og skisser et histogram basert på grupperingen. Kommenter formen på fordelingen.

c. De opprinnelige (usorterte) dataene betraktes som observasjoner av de stokastiske variablene, $X_1, X_2, \ldots, X_{130}$, som antas uavhengige og identisk fordelte med forventning og varians, $\mathrm{E}(X_i) = \mu$, $\mathrm{var}(X_i) = \sigma^2$, som begge antas ukjente. Tyder dataene på at μ er forskjellig fra 37,0? Sett opp null-hypotese og alternativ og en test for dette. Gjennomfør testen og undersøk om p-verdien er større eller mindre enn 0,002 (0,2%). Formuler en konklusjon.

d. Vi definerer en normaltemperatur som "uvanlig" dersom den er høyst lik 35,9 eller minst lik 37,5. La *p* betegne sannsynligheten for at en tilfeldig valgt person har uvanlig normaltemperatur. Gjør kort rede for hvorfor antall uvanlige normaltemperaturer i utvalget er binomisk fordelt. Bruk dette til å finne en forventningsrett estimator for $q = 100p$, dvs. prosenten av uvanlige normaltemperaturer i populasjonen. Beregn et estimat for q ut fra dataene i utvalget. Estimer også standardavviket for estimatoren din.

e. La intervallet (L, U) være et konfidensintervall for p med konfidensgrad $1-\alpha$. Forklar hvorfor konfidensintervallet $(100L,\ 100U)$ for q da også har samme konfidensgrad $1-\alpha$. Bruk dette til å beregne et konfidensintervall for q med konfidensgrad 0,90. Anslå også hvor mange observasjoner som må til for å få et 90% konfidensintervall for q med lengde høyst 2 prosentpoeng, dvs. slik at intervallet ser omtrent ut som $\hat{q} \pm 1$.

ENGLISH VERSION

Problem 1

In this problem you can assume that the probability of a girl (and the probability of a boy) is 1/2 in an arbitrary birth, and that the sex occurring in different births are independent. We are interested in the sex of the children in an arbitrary three-children family. Define the outcome *xyz* as the sex of the three children, where *x* denotes the sex of the oldest, *y* of the second oldest, and *z* the sex of the youngest child. For example, the outcome *GBG* means that the oldest and the youngest of the children are girls while the second oldest is a boy.

a. Write down the sample space which, in this case (for a three-children family), consists of 8 possible outcomes. Explain why it is reasonable to assume that the corresponding probability model is uniform.

b. Define the following events:

A = “The Family has 2 girls and one boy”
B = “The family has at least 2 girls”
C = “The oldest child is a girl”

i) Find P(A), P(B) og P(C).
ii) Find $P(B \cup C)$, $P(B \cap C)$ og $P(B|C)$.

c. **i)** About a certain three-children family we know that they have at least two girls. What is then the probability that the third child is a boy? Note that we do not know here if the third child is born as no. 1, 2 or 3.
ii) About another three-children family we know that the two oldest children are girls. What is then the probability that the third child is a boy?

d. A young couple want at least one child of each sex, and decide to continue to give birth to children until that target is acheived. For simplicity we assume that the first child they got was a girl. Let *X* denote the number of new attempts (child births) necessary until the first boy is born. After the aim (at least one of each sex) is achieved, they do not want any more children. Explain why it is reasonable to assume that *X* is geometrically distributed when we ignore complications as, e.g., possible infertility, twin births, and the fact that it is limited how many children a single couple can get in practice.

Let *Y* be the total number of children the couple ends up with in the end. I.e., $Y = X + 1$. Calculate the expectation and variance of *Y*, and the probability that *Y* becomes at least 7.

e. After having considered the matter, the couple decide that they, whatever happens, do not want more than 6 children. In other words, even if they end up with 6 girls, they will quit getting more children. Determine the distribution of *Y* (total number of children including the first born girl) in this case. I.e., fill in the relevant probabilities in the following table:

y	2	3	4	5	6
$P(Y = y)$	?	?	?	?	?

What is expected Y in this case?

Problem 2

The normal temperature in human beings varies somewhat from person to person. Until fairly recently it has been common to assume that the average normal temperature in the population of humans is $37,0$ degrees Celcius. Table 1 gives the results of measuring the normal temperature for a sample of $n = 130$ people where the data have been sorted in an increasing order. (The normal temperature for an individual is determined as the average of several measurements over time.) Note that the table consists of 10 rows and 13 columns.

Table 1 (Normal temperature for 130 people)

35.7	35.8	35.9	35,9	36,0	36,1	36,1	36,2	36,2	36,2	36,2	36,2	36,2
36,3	36,3	36,3	36,3	36,3	36,3	36,4	36,4	36,4	36,4	36,4	36,4	36,5
36,5	36,5	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6	36,6
36,6	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7	36,7
36,7	36,7	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8	36,8
36,8	36,8	36,8	36,8	36,9	36,9	36,9	36,9	36,9	36,9	36,9	36,9	36,9
36,9	36,9	36,9	37,0	37,0	37,0	37,0	37,0	37,0	37,0	37,0	37,0	37,0
37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1	37,1
37,1	37,1	37,1	37,1	37,1	37,2	37,2	37,2	37,2	37,2	37,2	37,2	37,3
37,3	37,3	37,3	37,3	37,3	37,4	37,4	37,4	37,4	37,5	37,7	37,8	38,2

Mean: 36,81 Standard deviation: 0,4074

a. Determine the median of the observations. Does it make any difference if the mean and standard deviation are calculated based on the sorted data in the table or based on the original (unsorted) data? Give a reason for your answer.

b. The observations are grouped in intervals shown in table 2. Note that the upper limit in each interval is inclusive while the lower limit is exclusive. This means, for example, that the observation, 36,1 belongs to the second interval and not the third.

Table 2

Interval	Frequency
35,5 - 35,8	2
35,8 - 36,1	5
36,1 - 36,4	18
36,4 - 36,7	29
36,7 - 37,0	37
37,0 - 37,3	31
37,3 - 37,6	?
37,6 - 37,9	?
37,9 - 38,2	?
Sum	130

Fill in the correct frequencies where it is a question mark in the table, and sketch an histogram based on the grouped data. Comment on the shape of the distribution.

c. The original (unsorted) data are considered to be observations of random variables, $X_1, X_2, \ldots, X_{130}$, which are assumed to be independent and identically distributed with expectation and variance, $\mathrm{E}(X_i) = \mu$, $\mathrm{var}(X_i) = \sigma^2$, both unknown. Is there evidence in the data that μ is different from 37,0? Write down the null hypothesis and alternative and establish a test for this. Perform the test and investigate whether the p-value is larger or smaller than 0,002 (0,2%). Formulate a conclusion.

d. We define the normal temperature as "unusual" if it is at most equal to 35,9 or at least equal to 37,5. Let *p* denote the probability that a randomly chosen person has unusual normal temperature. Explain briefly why the number of unusual normal temperatures in the sample is binomially distributed. Use this to find an unbiased estimator for $q = 100p$, i.e., the percentage of unusual normal temperatures in the population. Also estimate the standard deviation of your estimator.

e. Let the interval (L, U) be a confidence interval for *p* with the degree of confidence $1-\alpha$. Explain why the corresponding confidence interval $(100L,\ 100U)$ for *q* also has the same degree of confidence $1-\alpha$. Use this to calculate a 90% confidence interval for *q*. Estimate how many observations are needed to obtain a 90% confidence interval for *q* with length at most 2 percentage points, i.e., so that the resulting interval will look close to $\hat{q} \pm 1$.