

Institutt for samfunnsøkonomi

## Eksamensoppgave i SØK1004 – Statistikk for økonomer

**Faglig kontakt under eksamen: Isabel Hovdahl**

**Tlf.: 902 80 774**

**Eksamensdato:** 4. desember 2019

**Eksamenstid (fra-til):** 4 timer (15.00-19.00)

**Hjelpemiddelkode/Tillatte hjelpemidler: C**

Formelsamling:

Knut Sydsæter, Arne Strøm og Peter Berck (2006): Matematisk formelsamling for økonomer, 4utg. Gyldendal akademiske.

Knut Sydsæter, Arne Strøm, og Peter Berck (2005): Economists' mathematical manual, Berlin.

Kalkulator:

Casio fx-82ES PLUS, Casio fx-82EX Citizen SR-270x, SR-270X College eller HP 30S.

**Målform/språk:** Norsk og engelsk

**Antall sider bokmål (uten forside):** 3

**Antall sider engelsk (uten forside):** 3

**Antall sider vedlegg:** 10

**Informasjon om trykking av eksamensoppgave**

Originalen er:

1-sidig  2-sidig

sort/hvit  farger

skal ha flervalgskjema

**Kontrollert av:**

\_\_\_\_\_

Dato

\_\_\_\_\_

Sign

Eksamen består av 5 oppgaver med flere delsspørsmål som alle skal besvares.

**Oppgave 1 (20%)**

Finne det korrekte svaralternativet på hver oppgave. Utrekning skal ikke være med.

- a) La  $\bar{X}$  være gjennomsnittet av et tilfeldig utvalg av størrelse  $n = 4$  fra en normalfordelt populasjon med forventning  $\mu = 5$  og varians  $\sigma^2 = 16$ . Hva er sannsynligheten for at  $\bar{X}$  ligger mellom 6 og 8?
- A) 0.08      B) 0.43      C) 0.19      D) 0.24      E) 0.37
- b) Anta at sannsynligheten for diabetes-diagnose av en person med diabetes er 95%, og at sannsynligheten for diabetes-diagnose av en person uten diabetes er 5%. Gitt at 10% av populasjonen har diabetes, hva er sannsynligheten for at en person med diabetes-diagnose faktisk har diabetes?
- A) 0.65      B) 0.72      C) 0.80      D) 0.68      E) 0.22
- c) En håndballspiller tar 4 straffekast i løpet av en kamp. Det antas at spilleren har en sannsynlighet  $p = 0.8$  for å få mål på hvert straffekast og at straffekastene resulterer i mål eller ikke mål uavhengig av hverandre. Hva er sannsynlighet for at spilleren får mål på minst 3 av de 4 kastene?
- A) 0.51      B) 0.75      C) 0.82      D) 0.92      E) 0.24
- d) Tabellen under rapporterer informasjon for 531 personer som har vært utsatt for en sykkelulykke. For hver av personene er det registrert om de brukte sykkelhjelmer eller ikke ved ulykken, og om de pådro seg hodeskader eller ikke.

	Hodeskade	Ikke hodeskade	Totalt
Hjelm	30	83	113
Ikke hjelm	182	236	418
Totalt	212	319	531

Anta at det trekkes en person tilfeldig fra de 531, og at det viser seg at denne personen *ikke* fikk hodeskader ved ulykken. Hva er da sannsynligheten for at denne personen brukte hjelm?

- A) 0.73      B) 0.44      C) 0.60      D) 0.14      E) 0.26

- e) En tilfeldig variabel  $x$  med mulige verdier 1, 2, 3 og 4 har en sannsynlighetsfordeling gitt i tabellen nedenfor.

$x$	1	2	3	4
$P(x)$	0.4	0.2	0.1	0.3

Hva er variansen til  $x$ ?

- A) 2.30      B) 5.29      C) 1.61      D) 6.90      E) 1.27

### Oppgave 2 (10%)

Anta at  $X$  og  $Y$  er to avhengige variabler hvor  $X \sim N(1.5, 0.75)$  og  $Y \sim N(1.5, 0.75)$ . I tillegg får vi oppgitt at  $Cov(X, Y) = -0.5$ . Dersom  $Z = 2X - Y$ , finn fordelingen til  $Z$ .

### Oppgave 3 (30%)

En gruppe har undersøkt årsinntekten i et tilfeldig utvalg av 36 nyutdannede lærere. Vi får oppgitt følgende gjennomsnittlig årsinntekt og standardavvik i utvalget:

$$\bar{X} = 455\,000 \text{ kr}$$

$$s_X = 27\,000 \text{ kr}$$

- Finn et 95% konfidensintervall for  $\mu$  basert på gruppas undersøkelse. Gi en tolkning av konfidensintervallet.
- Hvor stort må utvalget være dersom forskjellen mellom  $\bar{X}$  og  $\mu$  skal være mindre enn 5000 kr med 95% sikkerhet?

I en eldre undersøkelse fant man at gjennomsnittlig årsinntekt blant nyutdannede lærere var 448 000 kr.

- Utfør en hypotesetest på et 5% signifikansnivå, og undersøk om det er grunnlag for å hevde at årsinntekten til nyutdannede lærere har økt.
- Gi en kort tolkning av signifikansnivået.

e) Hva er testens p-verdi? Gi en tolkning av p-verdien.

#### Oppgave 4 (20%)

Det hevdes at andelen som pådrar seg hodeskader i en sykkelulykke er lavere blant de som bruker hjelm enn blant de som ikke bruker hjelm. Bruk tabellen i oppgave 1d) til å teste denne påstanden på et 1% signifikansnivå.

#### Oppgave 5 (20%)

Doping er dessverre relativt utbredt i flere idretter. Betrakt 10 idrettsutøvere, og la  $X$  være mengden (målt i mg) av et dopingmiddel som en utøver får tilført, og la  $Y$  være mengden melkesyre (målt i mg) som ble målt etter en utholdenhetstest utført etter dopinginntaket. Tabellen under viser resultatene for de 10 idrettsutøverne.

$X$	10	20	20	25	30	40	50	50	60	70
$Y$	20	13	12	9.5	8	5	3.4	3	2	1.2

$$\sum_{i=1}^{10} (X_i - \bar{X}) = 375, \quad \sum_{i=1}^{10} (Y_i - \bar{Y}) = 77.1$$

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 3462.5, \quad \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 323.8, \quad \sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = -989.75$$

- Beregn korrelasjonskoeffisienten og begrunn om du synes det er en lineær sammenheng mellom  $X$  og  $Y$ .
- Anta at det er en lineær sammenheng mellom dopinginntak og melkesyre. Finn den estimerte regresjonslinjen ved bruk av dataene ovenfor. Illustrer utvalgets regresjonslinje i et diagram og gi en tolkning av linjen.
- Estimer mengden melkesyre for en person som har mottatt 15 mg av dopingmidlet.
- Beregn modellens forklaringskraft og kommenter den.

The exam consists of 5 questions. You must answer all of them.

**Question 1 (20%)**

Find the correct answer. Do not include your calculations.

- a) Let  $\bar{X}$  be the average of a random sample of size  $n = 4$  from a normally distributed population with expectation  $\mu = 5$  and variance  $\sigma^2 = 16$ . What is the probability that  $\bar{X}$  lies between 6 and 8?
- A) 0.08      B) 0.43      C) 0.19      D) 0.24      E) 0.37
- b) Assume that the probability of being diagnosed with diabetes if you have diabetes is 95%, and that the probability of being diagnosed with diabetes if you do not have diabetes is 5%. Given that 10% of the population has diabetes, what is the probability that a person diagnosed with diabetes does in fact have diabetes?
- A) 0.65      B) 0.72      C) 0.80      D) 0.68      E) 0.22
- c) A handball player makes 4 penalty throws during a match. It is assumed that the player has a probability of  $p = 0.8$  of scoring a point on each of the penalty throws and that the success of each penalty throw is independent of each other. What is the probability that the player will score a point on at least 3 out of the 4 throws?
- A) 0.51      B) 0.75      C) 0.82      D) 0.92      E) 0.24
- d) The table below reports information on 531 people that have been in a bicycle accident. For each of the person, it has been registered whether they suffered head trauma or not, and whether they wore a helmet or not.

	Head trauma	No head trauma	Total
Helmet	30	83	113
No helmet	182	236	418
Total	212	319	531

Assume that a person is randomly drawn from the sample of 531 people, and that this person did *not* experience head trauma from the bicycle accident. What is then the probability that this person wore a helmet?

- A) 0.73      B) 0.44      C) 0.60      D) 0.14      E) 0.26

- e) The table below reports the probability distribution of a random variable  $x$  that can take the values 1, 2, 3 and 4:

$x$	1	2	3	4
$P(x)$	0.4	0.2	0.1	0.3

What is the variance of  $x$ ?

- A) 2.30      B) 5.29      C) 1.61      D) 6.90      E) 1.27

### Question 2 (10%)

Assume that  $X$  and  $Y$  are two dependent variables where  $X \sim N(1.5, 0.75)$  and  $Y \sim N(1.5, 0.75)$ . In addition, we know that  $Cov(X, Y) = -0.5$ . If  $Z = 2X - Y$ , find the distribution of  $Z$ .

### Question 3 (30%)

A group of researchers have investigated the annual income of 36 newly qualified teachers. We are given the following information on the average and standard deviation of annual income in the sample:

$$\bar{X} = 455\,000 \text{ NOK}$$

$$s_X = 27\,000 \text{ NOK}$$

- Find a 95% confidence interval for  $\mu$  based on the sample. Give an interpretation of the confidence interval.
- How large must the sample size be if the difference between  $\bar{X}$  and  $\mu$  is to be less than 5000 NOK with 95% confidence?

In a previous investigation it was found that the average annual income of newly qualified teachers was 448 000 NOK.

- c) Using a 5% level of significance, test whether we can claim that the annual income of newly qualified teachers has risen.
- d) Give a short interpretation of the level of significance in hypothesis testing.
- e) What is the p-value of the test? Give an interpretation of the p-value.

**Question 4 (20%)**

It is claimed that the share that suffers head trauma from a bicycle accident is lower among those that wore a helmet on than those that did not have a helmet on. Use the table in question 1d) to test this claim using a 1% level of significance.

**Question 5 (20%)**

Unfortunately, the use of banned drugs is very common in many sports. Consider a sample of 10 athletes, where  $X$  is the amount of a banned drug (measured in mg) taken by the athlete, and  $Y$  is the amount of lactic acid (measured in mg) measured after a stamina test performed after the drug intake. The table beneath shows the results for the 10 athletes.

$X$	10	20	20	25	30	40	50	50	60	70
$Y$	20	13	12	9.5	8	5	3.4	3	2	1.2

$$\sum_{i=1}^{10} (X_i - \bar{X}) = 375, \quad \sum_{i=1}^{10} (Y_i - \bar{Y}) = 77.1$$

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 3462.5, \quad \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 323.8, \quad \sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = -989.75$$

- a) Find the correlation between  $X$  and  $Y$ , and comment on whether or not you think there is a linear relationship between the two variables.
- b) Assume that there is in fact a linear relationship between drug intake and lactic acid. Estimate the sample regression line using the data above. Illustrate the sample regression line in a plot and give an interpretation of the line.
- c) Predict the amount of lactic acid for a person that has taken 15 mg of the banned drug.
- d) Estimate the model's explanatory power and comment.

