

NAME:.....
 SPRING 2015-2016
 ECON 206
 ERDIL, AKDI

07.04.2016

MIDTERM EXAM I
 (TIME ALLOWED 120 MIN)
NO PARTIAL CREDIT

1. (25 pts) Let $X_1, X_2, \dots, X_n$ be a random sample from a normal population with mean μ and variance σ^2 . In order to estimate σ^2 , the following estimator

$$T_a = a \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

is proposed.

a) (15 pts) Find the value a which minimizes the mean squared error of T_a , (T_{MSE}).

$$\begin{aligned} MSE(T_a) &= Var(T_a) + [Bias(T_a)]^2 \\ E(T_a) &= E\left[a \sum_{i=1}^n (X_i - \bar{X}_n)^2\right] = E(a(n-1)S^2) = a(n-1)E(S^2) \\ &= a(n-1)\sigma^2 \\ Bias(T_a) &= E(T_a) - \sigma^2 = a(n-1)\sigma^2 - \sigma^2 \\ Var(T_a) &= \sigma^2 Var\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2\right) = \sigma^2 Var((n-1)S^2) \\ &= \sigma^2 (n-1)^2 Var(S^2) = 2\sigma^2 (n-1)\sigma^4 \end{aligned}$$

b) (10 pts) Calculate the MSE of T_{MSE}

$$\begin{aligned} T_{MSE} &= \frac{1}{n+1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n+1} S_n^2 \\ E(T_{MSE}) &= \frac{n-1}{n+1} \sigma^2 \\ Var(T_{MSE}) &= \left(\frac{n-1}{n+1}\right)^2 Var(S^2) = \frac{(n-1)^2}{(n+1)^2} \frac{2\sigma^4}{1} \\ Bias &= \left(\frac{n-1}{n+1} - 1\right) \sigma^2 = \frac{1-1-n+1}{n+1} = -\frac{2}{n+1} \\ MSE &= Var + Bias^2 \\ &= \frac{2\sigma^4(n-1)}{(n+1)^2} + \frac{4}{(n+1)^2} \end{aligned}$$

1) σ - continued.

$$\begin{aligned} \text{MSE}(\tau_0) &= 2\sigma^2(n-1)\sigma^4 + (\sigma(n-1)\sigma^2 - \sigma^2)^2 \\ &= 2\sigma^2(n-1)\sigma^4 + (\sigma(n-1) - 1)^2 \sigma^4 \\ &= 2\sigma^2(n-1)\sigma^4 + \sigma^2(n-1)\sigma^4 - 2\sigma(n-1)\sigma^4 + \sigma^4 \end{aligned}$$

$$\begin{aligned} \frac{\partial \text{MSE}(\tau_0)}{\partial \sigma} &= 4\sigma(n-1)\sigma^3 + 2\sigma(n-1)^2\sigma^3 - 2(n-1)\sigma^3 \\ &= 2(n-1)\sigma^3(2\sigma + \sigma(n-1) - 1) = 0 \end{aligned}$$

$$2\sigma + \sigma(n-1) - 1 = 0 \Rightarrow \sigma + \sigma n = 1$$

$$\boxed{\sigma = \frac{1}{n+1}}$$

2. (25 pts) Let $X_1, X_2, \dots, X_n$ be a random sample from

$$f(x; \theta) = (\theta + 1)x^\theta \quad 0 < x < 1$$

a) (10 pts) Find the method of moments estimator of θ ($\hat{\theta}$).

$$E(X) = \int_0^1 (\theta + 1)x^{\theta+1} dx = \frac{\theta+1}{\theta+2}$$

$$\frac{\hat{\theta}+1}{\hat{\theta}+2} = \bar{x} \Rightarrow \hat{\theta}+1 = \hat{\theta}\bar{x} + 2\bar{x}$$

$$\hat{\theta}(1-\bar{x}) = 2\bar{x}-1$$

$$\boxed{\hat{\theta} = \frac{2\bar{x}-1}{1-\bar{x}}}$$

b) (15 pts) Find the maximum likelihood estimator of θ ($\hat{\theta}$).

$$L = \prod_{i=1}^n (\theta+1)x_i^\theta = (\theta+1)^n \left(\prod_{i=1}^n x_i \right)^\theta$$

$$\ln(L) = n \ln(\theta+1) + \theta \sum_{i=1}^n \ln x_i$$

$$\frac{\partial \ln(L)}{\partial \theta} = \frac{n}{\theta+1} + \sum \ln x_i = 0$$

$$\frac{n}{\theta+1} = -\sum \ln x_i$$

$$\boxed{\hat{\theta} = \frac{n}{-\sum \ln x_i} - 1}$$

3) (20 pts)

a) (5 Points) Briefly state the central limit theorem.

b) (15 Points) Let $X_1, X_2, \dots, X_n$ be a random sample from a uniform distribution ($X_1, X_2, \dots, X_n$ are iid $U(0, 2)$) Find the probability that the sum of randomly selected 100 numbers from $U(0, 2)$ is less than 110 (Hint use central Limit Theorem and take $\sqrt{3} \cong 1.73$)

SOLUTIONS

(3) a) Briefly state the CLT

$X_1, X_2, \dots, X_n$ iid $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2 < \infty$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$$

b) $X_1, X_2, \dots, X_n$ iid $U(0, 2)$

$$f(x) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq 2 \\ 0, & \text{else.} \end{cases}$$

$$E(X) = \int_0^2 \frac{x}{2} dx = \frac{x^2}{4} \Big|_0^2 = 1$$

$$E(X^2) = \int_0^2 \frac{x^2}{2} dx = \frac{x^3}{6} \Big|_0^2 = \frac{8}{6}$$

$$\left. \begin{array}{l} E(X) = 1 \\ E(X^2) = \frac{8}{6} \end{array} \right\} \text{Var}(X) = \frac{8}{6} - 1 = \frac{2}{6} = \frac{1}{3}$$

$$P\left(\sum_{i=1}^n X_i \leq 110\right) = ?$$

$$E(\sum X_i) = 100, \text{Var}(X_i) = \frac{100}{3}$$

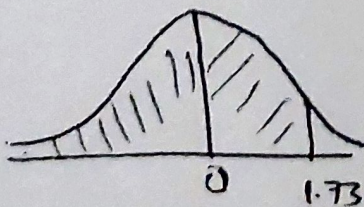
$$P\left(\sum_{i=1}^n X_i \leq 110\right) = P\left[\frac{\sum X_i - E(\sum X_i)}{\sqrt{\text{Var}(\sum X_i)}} \leq \frac{110 - 100}{\sqrt{\frac{100}{3}}}\right]$$

$$\approx P\left(Z \leq \frac{\sqrt{3}(110 - 100)}{10}\right) = P(Z \leq \sqrt{3})$$

$$= P(Z \leq 1.73) = 0.5 + P(0 < Z < 1.73)$$

$$= 0.5 + 0.4581$$

$$= \underline{\underline{0.9581}}$$



4) (30 pts) Let $X_1, X_2, \dots, X_n$ be a random sample from a normal population with mean μ and

variance σ^2 ($X_1, X_2, \dots, X_n \text{ iid } N(\mu, \sigma^2)$) Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

a) (5 Points) Show that $\bar{X}_n$ and S_n^2 are unbiased estimators for μ and σ^2 , respectively

b) (5 Points) Show that $\bar{X}_n$ and S_n^2 are consistent estimators for μ and σ^2 , respectively

c) (10 Points) Show that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{D} N(0,1)$ as $n \rightarrow \infty$

d) (10 Points) Show that $\bar{X}_n$ is the unique UMVUE estimator for μ .

② $X_1, X_2, \dots, X_n \text{ iid } N(\mu, \sigma^2)$

a) $E(\bar{X}_n) = \mu$, $Var(\bar{X}_n) = \sigma^2/n$

$$E(S_n^2) = \sigma^2 - \frac{n-1}{n} \sigma^2$$

So, $\bar{X}_n$ and S_n^2 are unbiased for μ & σ^2

b) $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \frac{\sigma^2}{n}$

$Bias_{\mu}(\bar{X}_n) = 0$ and $Var(\bar{X}_n) \rightarrow 0$ as $n \rightarrow \infty$

$\therefore \bar{X}_n \xrightarrow{P} \mu$ as $n \rightarrow \infty$

and $E(S_n^2) = \sigma^2$, $Var(S_n^2) = \frac{2\sigma^4}{n-1} \rightarrow 0$ as $n \rightarrow \infty$

$$\left(\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2 \right)$$

$Bias_{\sigma^2}(S_n^2) = 0$ and $Var(S_n^2) \rightarrow 0$

which implies that $S_n^2 \xrightarrow{P} \sigma^2$ as $n \rightarrow \infty$

Thus, $\bar{X}_n$ and S_n^2 are consistent for μ & σ^2 .

c) Note that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0,1)$

and $S_n^2 \xrightarrow{P} \sigma^2$, $\frac{S_n}{\sigma} \xrightarrow{P} 1$ as $n \rightarrow \infty$.

$\rightarrow$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} = \frac{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}{\frac{s_n}{\sigma}} \xrightarrow{D} N(0,1) \text{ as } n \rightarrow \infty$$

by Slutsky's th

d) Note that $E(\bar{X}_n) = \mu$ and $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$

$$\frac{d}{d\mu} E(\bar{X}_n) = 1$$

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$\ln f(x; \mu, \sigma^2) = -\frac{1}{2} \ln(\sigma^2) - \frac{1}{2} \ln(2\pi) - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{d}{d\mu} \ln f(x; \mu, \sigma^2) = 0 + 0 + \frac{2}{\sigma^2}(x-\mu) = \frac{2}{\sigma^2}x - \frac{2\mu}{\sigma^2}$$

$$= \frac{x}{\sigma^2} - \frac{\mu}{\sigma^2}$$

$$\frac{d^2}{d\mu^2} \ln f(x; \mu, \sigma^2) = -\frac{1}{\sigma^2}$$

so, $-n E\left(\frac{\partial^2}{\partial \mu^2} \ln f(x; \mu, \sigma^2)\right) = -n E\left(-\frac{1}{\sigma^2}\right) = \frac{n}{\sigma^2}$

By Cramér-Rao's inequality

$$\text{Var}(\bar{X}_n) \geq \frac{\left[\frac{d}{d\mu} E(\bar{X}_n)\right]^2}{-n E\left(\frac{\partial^2}{\partial \mu^2} \ln f(x; \mu, \sigma^2)\right)}$$

$$\begin{aligned} \frac{\sigma^2}{n} = \text{Var}(\bar{X}_n) &\geq \frac{\left[\frac{d}{d\mu} E(\bar{X}_n) \right]^2}{-nE\left(\frac{\partial^2}{\partial \mu^2} \ln f(X; \mu)\right)} \\ &= \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n} \end{aligned}$$

Since " $=$ " holds in Cramér-Rao's inequality for the estimator $\bar{X}_n$,

$\bar{X}_n$ is the unique UMVUE estimator for μ .

NAME:.....

Fall 2020-2021

ECON 206

ERDIL, AKDI

03.12.2020

MIDTERM
(TIME ALLOWED 120 MINS)
NO PARTIAL CREDIT

1. (25 pts.) Let $X_1, X_2, \dots, X_n$ be a random sample from a normal population with mean μ and variance σ^2 ($X_1, X_2, \dots, X_n \text{ iid } N(\mu, \sigma^2)$). In order to estimate σ^2 an estimator

$$T_c = c \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

is defined. Here, $\bar{X}_n$ denotes the sample mean.

- a) (15 pts.) Find the constant c which minimize the mean squared error of T_c
- b) (5 pts.) Calculate the mean and the variance of T_c
- c) (5 pts.) Calculate the mean squared error of T_c .
2. (25 pts.) Let $X_1, X_2, \dots, X_n$ be a random sample from a population with the following probability density function
- $$f(x; \theta) = \begin{cases} c e^{-(x-\theta)} & , \quad x \geq \theta \\ 0 & , \quad \text{elsewhere} . \end{cases}$$
- a) (3 pts.) Calculate the constant c
- b) (4 pts.) Find a sufficient estimator for θ (call T_n)
- c) (7 pts.) Based on the sufficient T_n , find an unbiased estimator for θ
- d) (7 pts.) Is the unbiased estimator you found in (c) consistent for θ
- e) (4 pts.) Find the unique UMVUE estimator for θ
3. (25 pts.) Let $X_1, X_2, \dots, X_n$ be a random sample from a population with the following probability density function

$$f(x; \beta) = \begin{cases} \frac{x}{\beta^2} e^{-\frac{x}{\beta}} & , \quad x \geq 0 \\ 0 & , \quad \text{elsewhere} . \end{cases}$$

- a) (15 pts.) Find the maximum likelihood estimator (MLE), $\hat{\beta}$ of β .
- b) (5 pts.) Is the estimator in part (a) an unbiased estimator for β . Why?
- c) (5 pts.) Is the estimator in part (a) a consistent estimator? Why?

4. (25 pts.) According to a study in an automotive plant, a trained worker should be able to assembly a certain type of component in 15 minutes.
- a) (13 pts.) Is this figure substantiated by the assembly times of 150 randomly selected chairs whose average assembly time is 15.4 minutes with standard deviation of 2.4 minutes? Use a two-sided test with 5% level of significance.
- b) (12 pts.) Calculate Type II error and the power of the test in part (a), given that true mean assembly time of these components is 15.5 minutes.

NAME:.....

SPRING 2020-2021

ECON 206

ERDIL, AKDI

03.05.2021

MIDTERM EXAM
(TIME ALLOWED 120 MIN)
NO PARTIAL CREDIT

1. (25 points) During the COVID-19 pandemic, the service times for patients coming through the emergency service of Ankara City Hospital are independent random variables with mean 1.5 minutes and variance 1.0. By using Central Limit theorem, approximate the probability that 100 patients can be served in less than 2 hours of total service. Comment on your finding.

2. (25 points) Let $X_1, X_2, \dots, X_n$ be a random sample from

$$f(x; \theta) = (\theta + 1)x^\theta \quad 0 < x < 1$$

a) (10 points) Find the method of moments estimator of θ ($\hat{\theta}$).

b) (15 points) Find the maximum likelihood estimator of θ ($\hat{\theta}$).

3. (25 points) Let $X_1, X_2, \dots, X_n$ be a random sample from a population with the following probability density function;

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & , \quad x > 0 \\ 0 & , \quad \text{elsewhere.} \end{cases}$$

In order to estimate the parameter θ , two estimators have been proposed;

$$T_1 = \bar{X}_n \text{ and } T_2 = nX_{(1)}$$

where $\bar{X}_n$ is the sample mean and $X_{(1)} = \min \{X_1, X_2, \dots, X_n\}$.

a) (10 points) Calculate the mean and the variance of the estimators T_1 and T_2

b) (10 points) Calculate the mean squared error of the estimators T_1 and T_2

c) (5 points) Show that $Var(T_1) \leq Var(T_2)$. Why?

4. (25 points) Let $X_1, X_2, \dots, X_n$ be a random sample from a population with the following probability density function;

$$f(x; \theta) = \begin{cases} \theta e^{-\theta x} & , \quad x > 0 \\ 0 & , \quad \text{elsewhere.} \end{cases}$$

a) (10 points) Find an unbiased estimator for the parameter θ (call T_n). Does T_n consistent for θ ? Show your work.

b) (15 points) Find a unique UMVUE estimator for θ (Call W_n) and calculate $Var(W_n)$

LE 7
MIDTERM EXAM
(03.05.2021)

① If we let γ_i denote the service time for the i th patient, then we want

$$P\left(\sum_{i=1}^{100} \gamma_i \leq 120\right) = P\left(\bar{\gamma} \leq \frac{120}{100}\right) = P(\bar{\gamma} \leq 1.20)$$

Because the sample size is large, CLT tells us that $\bar{\gamma}$ is approximately normally distributed with mean $\mu_{\bar{\gamma}} = \mu = 1.5$ and variance $\sigma_{\bar{\gamma}}^2 = \sigma^2/n = 1.0/100$. Thus, by using normal distribution table, we have

$$P(\bar{\gamma} \leq 1.20) = P\left(\frac{\bar{\gamma} - 1.50}{1/\sqrt{100}} \leq \frac{1.20 - 1.50}{1/\sqrt{100}}\right)$$

$$\approx P(Z \leq (1.2 - 1.5) \cdot 10) = P(Z \leq -3) = \boxed{0.0013}$$

Thus, the probability that 100 patients can be served in less than 2 hours is approximately 0.0013. This small probability indicates that it is virtually impossible to serve 100 patients in only 2 hours.

$$(2) \quad E(X) = \int_0^1 (\theta+1) x^\theta dx = \frac{\theta+1}{\theta+2}$$

$$\frac{\hat{\theta}+1}{\hat{\theta}+2} = \bar{x} \Rightarrow \hat{\theta}+1 = \hat{\theta}\bar{x} + 2\bar{x}$$

$$\hat{\theta}(1-\bar{x}) = 2\bar{x}-1$$

$$\boxed{\hat{\theta} = \frac{2\bar{x}-1}{1-\bar{x}}}$$

$$(b) \quad L = \prod_{i=1}^n (\theta+1) x_i^\theta = (\theta+1)^n \left(\prod_{i=1}^n x_i \right)^\theta$$

$$\ln L = n \ln(\theta+1) + \theta \sum \ln x_i$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\hat{\theta}+1} + \sum \ln x_i = 0$$

$$\frac{n}{\hat{\theta}+1} = -\sum \ln x_i$$

$$\boxed{\hat{\theta} = \frac{n}{-\sum \ln x_i} - 1}$$

Note that $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$.

$$(3) \quad f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & , x > 0 \\ 0 & , \text{else} \end{cases}$$

$$T_1 = \bar{X}_n \quad \text{and} \quad T_2 = nX_{(1)} = n \min\{X_1, \dots, X_n\}$$

a) It is clear that $E(X) = \theta$ and $\text{Var}(X) = \theta^2$ and thus

$$E(T_1) = \theta \quad \text{and} \quad \text{Var}(T_2) = \frac{\theta^2}{n}$$

In order to calculate the mean and Varian
 T_2 , We need to ~~calc~~ find the pdf of $X_{(1)}$

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(\min(X_1, \dots, X_n) > x)$$

$$= 1 - (e^{-x/\theta})^n = 1 - e^{-nx/\theta}$$

$$f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x) = \frac{n}{\theta} e^{-nx/\theta}, \quad x > 0$$

So $X_{(1)} \sim \exp(\theta/n)$ and thus

$$E(X_{(1)}) = \frac{\theta}{n} \quad \text{and} \quad \text{Var}(X_{(1)}) = \frac{\theta^2}{n^2}$$

$$E(T_2) = E(nX_{(1)}) = \theta \quad \text{and} \quad \text{Var}(T_2) =$$

$$\text{i.e. } E(T_1) = E(T_2) = \theta$$

$$\text{and } \text{Var}(T_1) = \frac{\theta^2}{n} \quad \text{and} \quad \text{Var}(T_2) = \theta^2$$

Note that T_1 is a consistent estimator

b. T_2 doesn't consistent for θ



$$(4) f(x|\theta) = \begin{cases} \theta e^{-\theta x} & , x > 0 \\ 0 & , \text{else} \end{cases}, X \sim \text{exp}(1/\theta)$$

$$b) \text{MSE}(T) = \text{Bias}_{\theta}^2(T) + \text{Var}_{\theta}(T)$$

Since both estimators unbiased for θ
 $(E(T_1) = E(T_2) = \theta)$, the MSE's are

$$\text{MSE}(T_1) = \text{Var}(T_1) = \frac{\theta^2}{n}$$

$$\text{MSE}(T_2) = \text{Var}(T_2) = \theta^2$$

Note that

$$\text{MSE}(T_1) \leq \text{MSE}(T_2) \text{ for } n > 1$$

$$c) \text{Var}(T_1) = \frac{\theta^2}{n} \leq \theta^2 = \text{Var}(T_2)$$

this is because T_1 is the unique UM
 estimator for θ . i.e. Among all unbiased
 estimators, T_1 has the smallest variance

$$(4) f(x|\theta) = \begin{cases} \theta e^{-\theta x} & x > 0 \\ 0 & \text{else} \end{cases}, X \sim \text{exp}(1/\theta)$$

$$\text{let } Y = \sum_{i=1}^n X_i \sim \text{Gamma}(n, 1/\theta)$$

$$f_Y(y) = \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y}, y > 0$$

$$E\left(\frac{1}{Y}\right) = \int_0^{\infty} \frac{1}{y} f_Y(y) dy = \frac{\theta^n}{\Gamma(n)} \int_0^{\infty} y^{n-2} e^{-\theta y} dy = \frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{\theta}{n-1}$$

and thus,

$$\boxed{T_n = \frac{n-1}{\sum X_i}} \text{ is unbiased for } \theta.$$

In order to check that it's consistent, we need to find the variance of T_n .

$$\begin{aligned} E\left(\frac{1}{Y^2}\right) &= \int_0^{\infty} \frac{1}{y^2} f_Y(y) dy = \frac{\theta^n}{\Gamma(n)} \int_0^{\infty} \frac{1}{y^2} y^{n-1} e^{-\theta y} dy \\ &= \frac{\theta^n}{\Gamma(n)} \int_0^{\infty} y^{n-3} e^{-\theta y} dy = \frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-2)}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)} \end{aligned}$$

$$\begin{aligned} \therefore \text{Var}\left(\frac{1}{Y}\right) &= E\left(\frac{1}{Y^2}\right) - \left(E\left(\frac{1}{Y}\right)\right)^2 = \frac{\theta^2}{(n-1)(n-2)} - \frac{\theta^2}{(n-1)^2} \\ &= \frac{(n-1)\theta^2 - (n-2)\theta^2}{(n-1)^2(n-2)} = \frac{\theta^2}{(n-1)^2(n-2)} \end{aligned}$$

and thus

$$\begin{aligned} \text{Var}(T_n) &= \text{Var}\left(\frac{n-1}{\sum X_i}\right) = \text{Var}\left(\frac{n-1}{Y}\right) = (n-1)^2 \text{Var}\left(\frac{1}{Y}\right) \\ &= (n-1)^2 \frac{\theta^2}{(n-1)^2(n-2)} = \frac{\theta^2}{n-2} \end{aligned}$$

T_n unbiased so $\text{Bias}(T_n) = 0$

$$\text{and } \text{Var}(T_n) = \frac{\theta^2}{n-2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Thus, T_n is consistent for θ .

b) Note that $\sum X_i$ is a sufficient estimator

for θ . $f(x; \theta) = \prod_{i=1}^n f(x_i; \theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} = g(T(x); \theta) \cdot h(x)$

By factorization thm, $\sum X_i$ is sufficient for θ .
(Moreover, it's complete, you do not need to check thm)

$T_n = \frac{n-1}{\sum X_i}$ is unbiased (from part (a))

By Rao-Blackwell thm,

$$E\left(\frac{n-1}{\sum X_i} \mid \sum X_i\right) = \frac{n-1}{\sum X_i} \text{ is the unique UMVUE}$$

estimator.

i.e. The UMVUE estimator is

$$W_n = \frac{n-1}{\sum X_i}$$

$$\text{Var}(W_n) = \frac{\theta^2}{n-2} \quad (\text{from part (a)})$$

NAME:.....
Middle East Technical University
Department of Economics
ECON 206
Erdil, Akdi

LET

05.05.2011

MIDTERM EXAM II
(Time Allowed: 120 minutes)
NO PARTIAL CREDIT

1. (15 pts.) Explain the concepts below:

- a) Type-I error.
- b) Type-II error.
- c) The power of the test. What are the factors that affect the power of a test?

Ⓐ Rejection of null hypothesis when it's true is called Type-I error. Its prob. is denoted by α which is also equal to size of critical region.

Ⓑ Fail to reject H_0 when it is false is called Type-II error, its prob. is denoted by β .

Ⓒ The power of a test is prob. that test will reject a false H_0 , not to commit Type-II error. Denoted by $1 - \beta$.

Power of test is affected by (ceteris paribus),

① Further the mean μ_1 from hypothesized mean μ_0 , greater the power of test. It means that we are more likely to detect large than small discrepancies from hypothesized mean.

② Smaller the significance level of test, smaller the power. Reducing prob. of Type-I error will increase prob. of Type-II error.

③ Larger the sample size, greater the power of test. More information is obtained from population.

⇒

④ Larger the population variance, lower the power of test. We are less likely to detect small departures from hypothesized mean when there is a greater variability in population.

2. (15 pts.) A factory produces lots of some merchandise A and the characteristics X of the elements of these lots has the distribution $N(\mu; \sigma^2)$, where the standard deviation σ is unknown. If the production process had not been significantly disturbed, σ should have remained constant. For this reason, σ plays the role of a measure of homogeneity or of a measure of accordance with the technological norms of the given production process. Samples were selected from the lots produced and the standard deviations of the samples are computed. We want to test the hypothesis that the population standard deviation is equal to 4 at 0.01 significance level. From a simple sample of size 15, sample standard deviation is observed as 4.5. Should we reject the hypothesis, that is, should we suppose that the observed value of standard deviation is significantly greater than the hypothetical value 4, so that this difference cannot be explained by random fluctuations if the null hypothesis is true?

$$H_0: \sigma^2 = 16$$

$$H_1: \sigma^2 > 16$$

$$\chi^2_{\alpha, n-1} = \chi^2_{0.01, 14} = 29.1413$$

$$\chi^2 = \frac{(n-1) s^2}{\sigma_0^2} \sim \chi^2_{\alpha, n-1}$$

$$\chi^2 = \frac{(15-1) \cdot (4.5)^2}{16} = 17.7188$$

$$\chi^2 < \chi^2_{0.01, 14} \Rightarrow \text{Not reject } H_0$$

The difference cannot be explained by random fluctuations.

$$H_0: \sigma = 4 \quad n = 15 \quad s = 4.5 \quad \alpha = 0.01$$

$$H_1: \sigma > 4$$

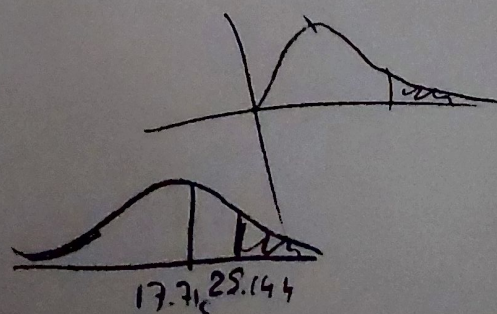
$$H_0: \sigma^2 = 16$$

$$H_1: \sigma^2 > 16$$

Failed to reject H_0 .

$$\chi^2_t = \frac{(n-1) \cdot s^2}{\sigma^2} = \frac{14 \cdot (4.5)^2}{16} = 17.71875$$

$$\chi^2_{\text{critic}} \sim \chi^2_{0.01, 14} = 29.141$$



3. (15 pts.) Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students. A random sample of 50 scores from each section produced the following results:

$$\text{Section I: } \bar{x}_1 = 77.01, s_1 = 10.32 \quad 50 = n_1, \alpha = 0.05$$

$$\text{Section II: } \bar{x}_2 = 72.22, s_2 = 11.02 \quad 50 = n_2$$

Construct a 95% confidence interval for the differences between means of these two sections.

$$\left[(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right] = 1 - \alpha$$

$$\left[(77.01 - 72.22) - 1.96 \sqrt{\frac{(10.32)^2}{50} + \frac{(11.02)^2}{50}} < \mu_1 - \mu_2 < \dots \right] = 0.95$$

$$\left[4.79 - 4.18 < \mu_1 - \mu_2 < 4.79 + 4.18 \right] = 0.95$$

$$\left[0.60 < \mu_1 - \mu_2 < 8.98 \right] = 0.95$$

$$\alpha = 0.05$$

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$t_{0.025}$

$$0.60 < \dots < 8.974885623$$

4. (20 pts.) If $H_0: \mu = 200$ is to be tested against $H_1: \mu < 200$ at 0.10 significance level based on a random sample of size n from a normal distribution where $\sigma = 15$, what is the smallest value for n that will make the power of the test equal to at least 0.75 when $\mu = 197$?

For $\alpha = 0.10$, $H_0: \mu = 200$ should be rejected if $\bar{X} < 200 - 1.28 \left(\frac{15}{\sqrt{n}} \right)$

$$1 - \beta = P \left[\bar{X} < 200 - 1.28 \left(\frac{15}{\sqrt{n}} \right) \mid \mu = 197 \right] = 0.75.$$

Thus;

$$P \left[\frac{200 - 1.28 \left(\frac{15}{\sqrt{n}} \right) - 197}{15 / \sqrt{n}} \right] = 0.75.$$

$$P(Z < 0.67) = 0.75 \Rightarrow$$

$$\frac{200 - 1.28 \left(\frac{15}{\sqrt{n}} \right) - 197}{15 / \sqrt{n}} = 0.67.$$

$$3 = \frac{10.05}{\sqrt{n}} + \frac{19.2}{\sqrt{n}}$$

$$\boxed{n \approx 35}$$

5. (15 pts.) A random sample of size $n=8$ is drawn from the uniform pdf, $f(y; \theta) = 1/\theta$, $0 \leq y \leq \theta$ for the purpose of testing

$$H_0: \theta = 2$$

$$H_1: \theta < 2$$

at 0.10 significance level. Suppose the decision rule is to be based on Y_8 , the largest order statistic. What would be the probability of committing a Type II error when $\theta = 1.7$?

If H_0 is true Y_8 should be close to values of largest order statistic that are much smaller than 2 would be evidence in favor $H_1: \theta < 2$. It follows that the decision rule should be

Reject $H_0: \theta = 2$ if $y_8 \leq c$.

where $P(Y_8 \leq c | H_0 \text{ is true}) = 0.10$.

$$f_{Y_8}(y; \theta = 2) = 8 \left(\frac{y}{2} \right)^7 \cdot \frac{1}{2} \quad 0 \leq y \leq 2.$$

thus, the constant c must satisfy.

$$\int_0^c 8 \left(\frac{y}{2} \right)^7 \cdot \frac{1}{2} dy = 0.10$$

$$\text{or } \left(\frac{c}{2} \right)^8 = 0.10 \Rightarrow \boxed{c = 1.50}$$

when $\theta = 1.7$, β is probability that Y_8 falls in non-rejection region.

$$\beta = P(Y_8 > 1.50 | \theta = 1.7)$$

$$= \int_{1.50}^{1.70} 8 \left(\frac{y}{1.7} \right)^7 \cdot \frac{1}{1.7} dy = 1 - \left(\frac{1.5}{1.7} \right)^8 = \boxed{0.63}$$

$$N \sim (0, 1)$$

6. (20 pts.) Suppose a sample of size n is to be drawn from a normal distribution where σ is known to be 14.3. How large does n have to be to guarantee that the length of the 95% confidence interval for μ will be less than 3.06?

The length of confidence interval is

$$2 z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \frac{2(1.96)(14.3)}{\sqrt{n}} = \frac{56.056}{\sqrt{n}}$$

$$\text{For } \frac{56.056}{\sqrt{n}} \leq 3.06 \Rightarrow n \geq \left(\frac{56.056}{3.06} \right)^2 = 335.58$$

$$\boxed{n \approx 336}$$

$$\bar{X} - \varepsilon < \mu < \bar{X} + \varepsilon$$

$$\varepsilon = \frac{z_{\alpha/2}}{1.96} \cdot \frac{\sigma}{\sqrt{n}} = \frac{z_{0.025}}{1.96} \cdot \frac{14.3}{\sqrt{n}}$$

$$\frac{56.056}{\sqrt{n}} \leq 3.06$$

$$\left(\frac{56.056}{3.06} \right)^2 < n$$

d

Questions

Read the questions carefully. Provide answers with complete sentences. Answers without explanation will NOT receive any point! Clearly show your work. NO CELL PHONES! NO Cheat Sheet! Use nothing more than a pencil and eraser. Only calculators allowed.

1. Using a random sample of $n=600$ forecasts were obtained, and it was determined that 382 students exceed the threshold to pass the exam.
 - a. Test the null hypothesis that 50% of the student population passes this exam. (Use two-sided alternative and set $\alpha=0.05$).
 - b. Determine the probability of type 2 error (Beta). ($H_a: p=0.55$)
2. An engineering research center claims that, through the use of a new computer control system, automobiles should achieve, on average, at least an additional 5 miles per gallon of gas. A random sample of 81 automobiles was used to evaluate this product. The sample mean increase in miles per gallon achieved was 4.5, and the sample standard deviation was 2.9 miles per gallon.
 - a. Test the hypothesis at the 5% level.
 - b. Find the p -value of this test, and interpret the results.
 - c. Explain how to use the p -value to test the hypothesis at the 5% level.
3. Consider a problem with the hypothesis test $H_0: \mu = 4$ vs. $H_1: \mu > 4$, and the following decision rule: Reject H_0 if

$$\frac{\bar{x} - 4}{0.08/\sqrt{25}} > 1.645$$

- a. Rewrite the decision rule explicitly in terms of $\bar{x}$.
 - b. Compute the probability of Type II error for $\mu = 4.06$.
4. A study was conducted by the Florida Game and Fish Commission to assess the amounts of chemical residues found in the brain tissue of brown pelicans. In a test for DDT, random samples of $n_1 = 10$ juveniles and $n_2 = 13$ nestlings produced the results shown in the accompanying table (measurements in parts per million).

<u>Juveniles</u>	<u>Nestlings</u>
$n_1 = 10$	$n_2 = 13$
$\bar{y}_1 = 0.041$	$\bar{y}_2 = 0.026$
$s_1 = 0.017$	$s_2 = 0.006$

 - a. Test the hypothesis that mean amounts of DDT found in juveniles and nestlings do not differ versus the alternative, that the juveniles have a larger mean. Use $\alpha = 0.05$.
 - b. Is there evidence that the mean for juveniles exceeds that for nestlings by more than 0.01 parts per million? Comment on the p -value.
5. The ages of random sample of 5 university professors are 39, 54, 61, 72, and 59. Using the information, find a 90% confidence interval for the population standard deviation. Assume that ages of university professors are normally distributed.

Answers

1. Let's first define the null hypothesis

a. $H_0: P = P_0 = 0.5$ and $H_a: P \neq 0.5$. Next define the Rejection Region

$$RR = \left(\left| \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \right| \geq Z_{\alpha/2} \right)$$

$\alpha = 0.05$ implies that $Z_{0.025} = 1.96$

$$RR = \left(\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \geq 1.96 \right) \text{ or } \left(\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \leq -1.96 \right)$$

Put it in another way,

$\hat{p}_x > 0.5 \pm 1.96 \sqrt{\frac{0.5(1-0.5)}{600}}$ Thus, if $\hat{p}_x > 0.54$ or $\hat{p}_x < 0.46$ then we reject the null hypothesis. Since the calculated ratio from the sample equals $382/600 = 0.637$, then the null hypothesis is rejected at 5% level.

Another way is to look at the Z-values. Our test statistic is defined as follows:

$$t = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{\frac{382}{600} - 0.5}{\sqrt{0.5 * 0.5 / 600}} = 6.695$$

Since $|Z| > Z_{0.25} = 1.96$ we reject H_0

- b. Type II error happens when we fail to reject the null hypothesis although the alternative is correct. In that case Beta is defined as follows:

$$\begin{aligned} \beta &= P(0.46 \leq \hat{p}_x \leq 0.54) = P \left[\frac{0.46 - p_a}{\sqrt{p_a(1-p_a)/n}} \leq Z \leq \frac{0.54 - p_a}{\sqrt{p_a(1-p_a)/n}} \right] \\ &= P \left[\frac{0.46 - 0.55}{\sqrt{0.55(1-0.55)/600}} \leq Z \leq \frac{0.54 - 0.55}{\sqrt{0.55(1-0.55)/600}} \right] \\ &= P(-4.43 \leq Z \leq -0.49) = 0.3121 \end{aligned}$$

2. Here is a step-by-step solution.

a. $H_0: \mu = 5$ and $H_1: \mu < 5$. We reject H_0 if $Z < -Z_{0.05} = -1.645$

$Z_{\text{calculated}} = \frac{4.5-5}{2.9/\sqrt{81}} = -1.55$, therefore, we fail to reject H_0 at the 5% level. We conclude that through the use of a new computer control system, automobiles should achieve, on average, at least an additional 5 miles per gallon of gas.

b. p-value = $P(Z < -1.55) = 0.50 - 0.4394 = 0.0606$.

c. Since p-value = $0.0606 > \alpha = 0.05$, we fail to reject H_0 at the 5% level. In fact, as long as our alpha is below the p-value ($\alpha < 0.0606$), we will fail reject the null hypothesis. However, if we choose a greater alpha, such as $\alpha = 0.1$, then we will reject the null hypothesis. (Draw Graphs)

3. First rewrite the rejection rule

$$a. \frac{\bar{x} - 4}{0.08/\sqrt{25}} > 1.645 \Rightarrow \bar{x} > 4 + 1.645(0.08 / \sqrt{25}) = 4.026 \Rightarrow \text{Reject } H_0 \text{ if } \bar{x} > 4.026.$$

$$b. \beta = P(\bar{x} \leq 4.026 | \mu = 4.06) = P \left\{ Z \leq \frac{4.026 - 4.06}{0.08/\sqrt{25}} \right\} = P(Z \leq -2.125) = 0.50 - 0.4832 = 0.0168$$

4. Solution:

$$a. H_0: \mu_1 - \mu_2 = 0 \text{ vs } H_a: \mu_1 - \mu_2 > 0.$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{9(0.017)^2 + 12(0.006)^2}{21} = 0.00014443$$

$$s_p = 0.012017$$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2 \left[\left(\frac{1}{n_1} \right) + \left(\frac{1}{n_2} \right) \right]}} = \frac{0.041 - 0.026}{\sqrt{0.00014443 \left(\frac{1}{10} + \frac{1}{13} \right)}} = 2.97$$

The rejection region with $\alpha = 0.05$ and 21 degrees of freedom is $t > 1.721$ and the null hypothesis is rejected.

$$\text{b. } H_0 : \mu_1 - \mu_2 \leq 0.01 \quad \text{v.s.} \quad H_a : \mu_1 - \mu_2 > 0.01$$

$$t = \frac{(0.041 - 0.026) - 0.01}{\sqrt{0.00014443 \left(\frac{1}{10} + \frac{1}{13} \right)}} = 0.989 \quad \text{from table, the p-value is greater than 0.10. It is}$$

actually better to indicate that the p-value is between 0.1 and 0.25.

5. Solution

$$\text{Calculate, } \sum y_i = 285, \quad \sum y_i^2 = 16823 \quad \sum y_i^2 - ((\sum y_i)^2/n) = 578$$

Then the 90 % confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{0.05}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{0.95}} \quad \text{or} \quad \frac{578}{9.48773} < \sigma^2 < \frac{578}{0.710721}$$

$$\text{or } 60.92 < \sigma^2 < 813.26$$

which means that the 90% confidence interval for σ is $7.81 < \sigma < 28.52$

NAME:
Middle East Technical University
Department of Economics
ECON 206
Akdi, Erdil

15.05.2006

FINAL EXAM
(Time Allowed: 150 minutes)
NO PARTIAL CREDIT

1. (30 pts.) The following table shows the average rates of growth of gross domestic product, GDP, and employment E, for twenty-five OECD countries for the period 1988-1997.

Employment			GDP		
Australia	1.68	3.04	Korea	2.57	7.73
Austria	0.65	2.55	Luxembourg	3.02	5.64
Belgium	0.34	2.16	Netherlands	1.88	2.86
Canada	1.17	2.03	New Zealand	0.91	2.01
Denmark	0.02	2.02	Norway	0.36	2.98
Finland	-1.06	1.78	Portugal	0.33	2.79
France	0.28	2.08	Spain	0.89	2.60
Germany	0.08	2.71	Sweden	-0.94	1.17
Greece	0.87	2.08	Switzerland	0.79	1.15
Iceland	-0.13	1.54	Turkey	2.02	4.18
Ireland	2.16	6.40	United Kingdom	0.66	1.97
Italy	-0.30	1.68	United States	1.53	2.46
Japan	1.06	2.81			

Consider the following model

$$E_i = \beta_0 + \beta_1 G_i + u_i \quad i = 1, \dots, 25.$$

$$y_i = E_i$$

$$x_i = G_i$$

where E denotes the average growth of employment and G is the average growth of gross domestic product.

You are provided the following information from the sample:

$$\sum_{i=1}^{25} E_i = 20.84, \quad \sum_{i=1}^{25} G_i = 70.42, \quad \sum_{i=1}^{25} E_i G_i = 88.46, \quad \sum_{i=1}^{25} E_i^2 = 42.07$$

$$\sum_{i=1}^{25} G_i^2 = 259.13, \quad \sum_{i=1}^{25} e_i g_i = 29.76, \quad \sum_{i=1}^{25} e_i^2 = 24.71, \quad \sum_{i=1}^{125} g_i^2 = 60.77$$

where $e = (E - \bar{E})$ and $gdp = (G - \bar{G})$.



a) Obtain the least squares estimates of β_0 and β_1 .

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n e_i g_i}{\sum_{i=1}^n g_i^2} = \frac{29.76}{60.77} = 0.49, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 0.83 - (0.49)(2.82) = -0.55$$

The OLS fit is: $\hat{Y}_i = -0.55 + 0.49 X_i$

For the rest of the solutions, let us calculate the ANOVA table.

$$SSR = (\hat{\beta}_0, \hat{\beta}_1) \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix} - n(\bar{Y})^2 = (-0.55, 0.49) \begin{pmatrix} 20.84 \\ 88.46 \end{pmatrix} - 25(0.83)^2 = 14.66$$

$$SST = \sum (Y_i - \bar{Y})^2 = \sum e_i^2 = 24.71 \text{ and thus } SSE = SST - SSR = 24.71 - 14.66 = 10.05$$

The ANOVA Table

SoV	df	SS	MS	F
Regression	1	14.66	14.66	33.55
Error	23	10.05	0.437	
Total	24	24.71		

b) Calculate the coefficient of determination (R^2), simple correlation coefficient between E_i and $\hat{E}_i$ ($r_{\hat{Y}Y}$), and simple correlation coefficient between G_i and E_i (r_{XY}) then interpret them.

$$R^2 = \frac{SSR}{SST} = \frac{14.66}{24.71} = 0.59, \quad r_{\hat{Y}Y} = \sqrt{R^2} = \sqrt{0.59} = 0.77, \quad r_{XY} = 0.77$$

c) Obtain the least squares estimate of the variance of the model.

From the ANOVA table, $\hat{\sigma}^2 = MSE = 0.437$

d) Obtain the least squares estimates of the variances of the coefficients.

$$S^2(\hat{\beta}_0) = MSE \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} = \frac{(0.437)(259.13)}{25(60.77)} = 0.075$$

$$S^2(\hat{\beta}_1) = \frac{MSE}{\sum (X_i - \bar{X})^2} = \frac{(0.437)}{(60.77)} = 0.0072$$

e) Construct a 95% confidence interval for β_0 .

$$\hat{\beta}_0 \pm S(\hat{\beta}_0) t(23, 1 - \alpha/2) \Leftrightarrow (-0.55 \pm \sqrt{0.075} (2.069)) \Leftrightarrow (-1.109, 0.0086)$$

0.02

f) Evaluate whether the average growth of gross domestic product has both statistically and economically significant impact on average growth of employment.

Economically significant, a positive impact is expected.

$$H_0 : \beta_1 = 0 \text{ vs. } H_a : \beta_1 \neq 0.$$

$t = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} = \frac{0.49}{\sqrt{0.0072}} = 5.76$. Since $|t| > t(\text{table})$, we reject the null hypothesis. That is, statistically significant.

g) Test whether the regression model as a whole is significant.

Test, $H_0 : \beta_0 = \beta_1 = 0$. Reject the null if $F > F^\alpha(1, 23)$. From the ANOVA table $F = 33.55$ and thus, reject the null.

h) Test if $\beta_0 + \beta_1 = 1$.

$$\text{Note that } S^2(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X} S^2(\hat{\beta}_1) = (-2.82)(0.0072) = -0.02$$

$$S^2(\hat{\beta}_0 + \hat{\beta}_1) = S^2(\hat{\beta}_0) + S^2(\hat{\beta}_1) + 2S^2(\hat{\beta}_0, \hat{\beta}_1) = 0.075 + 0.0072 - 2(0.02) = 0.0422$$

$$t = \frac{(\hat{\beta}_0 + \hat{\beta}_1) - 1}{\sqrt{S^2(\hat{\beta}_0 + \hat{\beta}_1)}} = \frac{(-0.55 + 0.49) - 1}{\sqrt{0.0422}} = \frac{-1.06}{0.205} = -5.17$$

Since, $|t| > t(\text{table}) = 2.069$, we reject the null.

i) Interpret the model you have estimated and explain in detail whether the model should be reformulated.

Both coefficients are statistically and economically significant. Therefore, there is no need for reformulation.

2. (20 pts.) Consider a hypothetical country called non-corruption-land economy (Yolsuzluğa damardan girdik ekonomisi!). The following alternative regressions of GNP (G) on money stock (M) for the period 1990-2003 are provided. Figures in parentheses are respective standard errors for the coefficients and the geometric mean of the dependent variable is $\bar{G}_G = 1839.77$.

$$(I) \quad G_t = -44.06 + 1.59M_t \quad SSR = 75\,375 \quad \sigma = 79.254 \quad R^2 = 0.991$$

$$(61.01) \quad (0.04)$$

$$(II) \quad \ln G_t = 0.25 + 1.03 \ln M_t \quad SSR = 0.015 \quad \sigma = 0.035 \quad R^2 = 0.993$$

$$(0.18) \quad (0.02)$$

$$(III) \quad \ln G_t = 6.48 + 0.0008M_t \quad SSR = 0.072 \quad \sigma = 0.078 \quad R^2 = 0.966$$

$$(0.06) \quad (0.00004)$$

$$(IV) \quad G_t = -11884.97 + 1958.19 \ln M_t \quad SSR = 264\,854 \quad \sigma = 148.564 \quad R^2 = 0.967$$

$$(744.16) \quad (104.94)$$

a) By using the above information only for model (I),

i) Predict the level of G for the year 2004 when the money stock is 2 200 \$ billion.

The predicted model is given by

$$\hat{Y}_t = -44.06 + 1.59 X_t$$

Put $X_0 = 2200$ to get the predicted value as

$$\hat{G}_0 = -44.06 + 1.59(2200) = 3453.94$$

ii) Obtain the least squares estimate of the prediction error variance.

$$S_{ep}^2 = MSE \left[1 + \frac{1}{n} + \left(x_p^2 / \sum x_i^2 \right) \right] = 6281.2 \left[1 + \frac{1}{14} + (853275 / 3125087) \right] = 8444.624$$

iii) Construct a 95% confidence interval for Y_{2004} when $P_{2004} = 2\,250$ \$ billion.

$$\hat{Y}_p \pm S(\hat{Y}_p) t(n-2, 1-\alpha/2), \quad \hat{Y}_p = 3453.94$$

$$3453.94 \pm \sqrt{8444.624} (2.179) \Leftrightarrow (3253.701, 3654.178)$$

iv) Test whether prediction error is statistically significant when $Q_{2000} = 103$.

$$H_0 : E(ep) = 0 \text{ vs } H_a : E(ep) \neq 0$$

$$t = \frac{ep}{S(ep)} = \frac{3500 - 3453.94}{\sqrt{8444.624}} = 0.50. \text{ Since, } |t| < t(12, 1 - \alpha/2) \text{ we fail to reject the null.}$$

Prediction error is statistically insignificant

b) Which of the above equations can be chosen as the best demand specification? Show your calculations explicitly.

$$\text{Compare the Models (I) and (IV) : } R_I^2 > R_{IV}^2 \Rightarrow \text{Choose Model (I)}$$

$$\text{Compare the Models (II) and (III) : } R_{II}^2 > R_{III}^2 \Rightarrow \text{Choose Model (II)}$$

Then compare models (I) and (II)

$$S_L = 75375, S_{LL} = 0.015, k = 1/\bar{G}_G = 1/1839.77 = 0.00054$$

$$\phi = k^2 S_L / S_{LL} = 1.457 > 1$$

Choose log-linear specification given by (II)

3. (10 pts.) If $H_0: \mu = 220$ is tested against $H_1: \mu < 220$ at the 0.10 level of significance based on a random sample of size n from a normal distribution with $\sigma = 17.0$, what is the smallest value for n that will make the power of the test equal to at least 0.75 when $\mu = 207$.

$$\text{Reject the null hypothesis if } Z = \frac{\sqrt{n}(\bar{X} - 220)}{\sigma} < -z_\alpha = -1.28$$

$$\begin{aligned} \beta &= P(\text{Reject } H_0 | \mu = 207) = P(Z < -1.28 | \mu = 207) = P\left(\frac{\sqrt{n}(\bar{X} - 220)}{\sigma} < -1.28 | \mu = 207\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X} - 207 + 207 - 220)}{17} < -1.28 | \mu = 207\right) = P\left(\frac{\sqrt{n}(\bar{X} - 207)}{17} < -1.28 + \frac{13\sqrt{n}}{17}\right) \\ P\left(Z < -1.28 + \frac{13\sqrt{n}}{17}\right) &= 0.75 \end{aligned}$$

$$\text{Thus, from the normal table, } -1.28 + \frac{13\sqrt{n}}{17} = 0.67 \Rightarrow n = 6.5025, n \in N \Rightarrow n = 7$$

4. Suppose we run a simple regression of Y on X (That is, $Y_t = \beta_0 + \beta_1 X_t + e_t, t = 1, 2, \dots, n$) and obtained the following ANOVA table.

SoV	df	SS	MS	F
Regression	1	?	180	?
Error	?	72	?	
Total	9	?		

a) Complete the cells with “?” in the table.

b) Calculate R^2 and $\hat{\sigma}^2$.

c) Suppose the estimated value of β_1 is 2 and $\sum_{i=1}^n (X_i - \bar{X})^2 = 9$. Construct a 95% confidence interval for β_1 .

d) Test whether $\beta_1 = 1.5$ against the alternative $\beta_1 > 1.5$. Use $\alpha = 0.05$

Solution:

a)

SoV	df	SS	MS	F
Regression	1	180	180	20
Error	8	72	9	
Total	9	252		

b) $R^2 = \frac{SSR}{SSE} = \frac{180}{252} = 0.7142$ and $\hat{\sigma}^2 = MSE = 9$

c) A 95% confidence interval for β_1 is $\hat{\beta}_1 \pm s(\hat{\beta}_1)t_{n-2}(1-\alpha/2)$ where

$$s(\hat{\beta}_1) = \sqrt{MSE / \sum_{i=1}^n (X_i - \bar{X})^2} = \sqrt{9/9} = 1 \text{ and } t_8(0.975) = 2.306.$$

Therefore, a 95% confidence interval for β_1 is $2 \pm 1(2.306)$. That is, $(-0.306, 4.306)$.

d) $H_0: \beta_1 = 1.5$ vs $H_a: \beta_1 > 1.5$ and reject H_0 if $t_c > t_8(1-\alpha)$ where

$$t_c = \frac{\hat{\beta}_1 - 1.5}{s(\hat{\beta}_1)} = \frac{2 - 1.5}{1} = 0.5 \text{ and } t_8(0.95) = 1.8595.$$

Conclusion: Since $t_c < t_8(0.95)$, we fail to reject H_0 .

5. Consider a regression model without intercept. $Y_t = \beta X_t + e_t, t = 1, 2, \dots, n$. Assume that the errors are independent with mean zero and variance $\sigma^2 X_t^2$.

a) Write the OLS estimator of β (say $\hat{\beta}$). Calculate the mean and the variance of $\hat{\beta}$.

b) Another estimator of β is given by $\frac{1}{n} \sum_{i=1}^n (Y_i / X_i)$, (say $\hat{\beta}^*$). Calculate the mean and the variance of $\hat{\beta}^*$ and compare its variance with the variance of the OLS estimator.

Solution:

a) The OLS estimator of β is

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \text{ with mean } E(\hat{\beta}) = \frac{\sum_{i=1}^n X_i E(Y_i)}{\sum_{i=1}^n X_i^2} = \frac{\sum_{i=1}^n X_i \beta X_i}{\sum_{i=1}^n X_i^2} = \beta \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2} = \beta$$

and the variance

$$Var(\hat{\beta}) = \frac{\sum_{i=1}^n X_i^2 Var(Y_i)}{\left(\sum_{i=1}^n X_i^2\right)^2} = \frac{\sum_{i=1}^n X_i^2 \sigma^2 X_i^2}{\left(\sum_{i=1}^n X_i^2\right)^2} = \frac{\sigma^2 \sum_{i=1}^n X_i^4}{\left(\sum_{i=1}^n X_i^2\right)^2}$$

$$b) E(\hat{\beta}^*) = E\left(\frac{1}{n} \sum_{i=1}^n (Y_i / X_i)\right) = \frac{1}{n} \sum_{i=1}^n E(Y_i) / X_i = \frac{1}{n} \sum_{i=1}^n \beta X_i / X_i = \beta$$

and the variance

$$Var(\hat{\beta}^*) = Var\left(\frac{1}{n} \sum_{i=1}^n (Y_i / X_i)\right) = \frac{1}{n^2} \sum_{i=1}^n Var(Y_i) / X_i^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 X_i^2 / X_i^2 = \frac{\sigma^2}{n}$$

Now, comparison: Let $\underline{a} = (1, 1, 1, \dots, 1)'$ and $\underline{b} = (X_1^2, X_2^2, \dots, X_n^2)'$. Remember the Cauchy-Schwarz inequality:

$$\left(\underline{a}' \underline{b}\right)^2 \leq \left(\underline{a}' \underline{a}\right) \left(\underline{b}' \underline{b}\right).$$

Note that $\underline{a}' \underline{b} = \sum_{i=1}^n X_i^2$, $\underline{a}' \underline{a} = n$ and $\underline{b}' \underline{b} = \sum_{i=1}^n X_i^4$. Therefore,

$$\left(\underline{a}' \underline{b}\right)^2 = \left(\sum_{i=1}^n X_i^2\right)^2 \leq \left(\underline{a}' \underline{a}\right) \left(\underline{b}' \underline{b}\right) = n \left(\sum_{i=1}^n X_i^4\right). \text{ That is,}$$

$$\left(\sum_{i=1}^n X_i^2\right)^2 \leq n \left(\sum_{i=1}^n X_i^4\right) \Leftrightarrow \frac{\sigma^2}{n} \leq \frac{\sigma^2 \left(\sum_{i=1}^n X_i^4\right)}{\left(\sum_{i=1}^n X_i^2\right)^2} \Leftrightarrow \text{Var}(\hat{\beta}^*) \leq \text{Var}(\hat{\beta}).$$

That is, the OLS estimator has larger variance than the other.

NAME:.....
 SPRING 2016-2017
 ECON 206
 ERDIL, AKDI

KEY

08.06.2017

FINAL EXAM
 (TIME ALLOWED 130 MIN)
NO PARTIAL CREDIT

1. (50pts.) Consider the following model for the relation between average GDP and average employment growth for 25 OECD countries for the period 1988-2012.

$$L_t = \beta_0 + \beta_1 G_t + u_t \quad t = 1988, \dots, 2012.$$

where L denotes the average rate of employment growth and G is the average rate of GDP growth.

You are provided the following information from the sample:

$$\sum_{t=1988}^{2012} L_t = 20.84, \quad \sum_{t=1988}^{2012} G_t = 70.42, \quad \sum_{t=1988}^{2012} G_t L_t = 88.4636, \quad \sum_{t=1988}^{2012} L_t^2 = 42.0742$$

$$\sum_{t=1988}^{2012} G_t^2 = 259.1294, \quad \sum_{t=1988}^{2012} g_t l_t = 29.7615, \quad \sum_{t=1988}^{2012} g_t^2 = 60.7703, \quad \sum_{t=1988}^{2012} l_t^2 = 24.7020$$

where $g = (G - \bar{G})$ and $l = (L - \bar{L})$.

LET $\alpha = 0.05$ FOR ALL TESTS OF SIGNIFICANCE UNLESS OTHERWISE STATED.

a) (4 pts.) Obtain the least squares estimates of β_0 and β_1 .

$$\hat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2} = 28.7615 / 60.7703$$

$$\boxed{\hat{\beta}_1 = 0.4857}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \left(\frac{20.84}{25} \right) - 0.4857 \left(\frac{70.42}{25} \right)$$

$$= 0.8336 - (0.4857)(2.8168)$$

$$= 0.8336 - 1.3754$$

$$\boxed{\hat{\beta}_0 = -0.5458}$$

$$\hat{L}_t = -0.5458 + 0.4857 G_t, \quad t = 1988, \dots, 2012.$$

- b) (3pts.) Calculate the coefficient of determination (R^2), simple correlation coefficient between L and $\hat{L}$ ($r_{\hat{L}}$), and simple correlation coefficient between G and L (r_{GL}) then interpret them.

$$R^2 = \frac{(\sum \hat{L}_t \cdot Y_t)^2}{\sum \hat{L}_t^2 \sum Y_t^2} = \frac{(28.7615)^2}{(60.7703)(24.72)} = \boxed{0.53}$$

$$r_{\hat{L}Y} = r_{XY} = \sqrt{0.53} = \boxed{0.7681}$$

- c) (3pts.) Obtain the least squares estimate of the variance of the model.

$$\begin{aligned} \sum e_t^2 &= \sum Y_t^2 - \hat{\beta}_0 \sum Y_t - \hat{\beta}_1 \sum X_t Y_t \\ &= 42.0742 - (-0.3458)(20.84) - (0.4137)(38.4136) \\ &= 42.0742 - (-11.3745) - 15.8916 \\ &= 10.1281 \end{aligned}$$

$$\sigma^2 = \frac{\sum e_t^2}{T-2} = \frac{10.1281}{25-2} = \boxed{0.4404}$$

- d) (3pts.) Obtain the least squares estimates of the variances of the coefficients.

$$\begin{aligned} \frac{\sigma^2}{\hat{\beta}_0} &= \sigma^2 \sum X_t^2 / T \sum Y_t^2 \\ &= (0.4404)(253.1284) / (25)(60.7703) \\ &= 114.1206 / 1518.2575 = \boxed{0.0751} \end{aligned}$$

$$\begin{aligned} \frac{\sigma^2}{\hat{\beta}_1} &= \sigma^2 / \sum X_t^2 = 0.4404 / 60.7703 \\ &= \boxed{0.0072} \end{aligned}$$

e) (3pts.) Construct a 95% confidence interval for β_0 .

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{0.0757} = 0.2740, \quad t_{23, 0.025} = 2.068$$

$$P(\hat{\beta}_0 - \hat{\sigma}_{\hat{\beta}_0} \cdot t_{T-2, \alpha/2} \leq \beta_0 \leq \hat{\beta}_0 + \hat{\sigma}_{\hat{\beta}_0} \cdot t_{T-2, \alpha/2}) = 0.95$$

$$P(-0.7458 - (0.2740)(2.068) \leq \beta_0 \leq -0.7458 + (0.2740)(2.068)) = 0.95$$

$$P[-1.4427 \leq \beta_0 \leq -0.0211] = 0.95$$

f) (3pts.) Evaluate whether slope coefficient is both statistically and economically significant.

$$H_0: \beta_1 = 0 \quad \hat{\sigma}_{\hat{\beta}_1} = \sqrt{0.0092} = 0.0848$$

$$H_1: \beta_1 \neq 0 \quad t_{23, 0.025} = 2.068$$

$$t = \hat{\beta}_1 / \hat{\sigma}_{\hat{\beta}_1} = 0.4857 / 0.0848 = \boxed{5.7680}$$

$t > t_{\alpha/2, 23} \Rightarrow \text{Reject } H_0$

$\hat{\beta}_1$ is statistically significant. It is also economically significant since there is a positive relation b/w GDP growth & employment growth.

g) (3pts.) Test whether the regression model as a whole is significant.

$$H_0: \beta_0 = \beta_1 = 0$$

$$Q = \left(\frac{R^2}{1-R^2} \right) \left(\frac{T-k-1}{p} \right) = \left(\frac{0.55}{1-0.55} \right) \left(\frac{23-2}{2} \right)$$

$$Q = (1.4080)(11.5) = \boxed{16.1920}$$

$Q > p_{\alpha, 23, 2} \Rightarrow \text{Reject } H_0$

The model is statistically significant as a whole.

h) (3pts.) Test if $\beta_0 + \beta_1 = 1$.

$$t_{23, 0.025} = 2.065$$

$$H_0: \beta_0 + \beta_1 = 1$$

$$H_1: \beta_0 + \beta_1 \neq 1$$

$$t_{\hat{\beta}_0 + \hat{\beta}_1} = \frac{(\hat{\beta}_0 + \hat{\beta}_1) - 1}{\sigma_{\hat{\beta}_0 + \hat{\beta}_1}}$$

$$\sigma_{\hat{\beta}_0 + \hat{\beta}_1}^2 = \sigma_{\hat{\beta}_0}^2 + \sigma_{\hat{\beta}_1}^2 - 2 \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$= 0.0751 + 0.0092 - 0.0823 \Rightarrow \sigma_{\hat{\beta}_0 + \hat{\beta}_1} = 0.2865$$

$$t_{\hat{\beta}_0 + \hat{\beta}_1} = \frac{(0.4857 - 0.5458) - 1}{0.2865} = \frac{-1.0561}{0.2865} = -3.687$$

$$|t_{\hat{\beta}_0 + \hat{\beta}_1}| > t_{23, 0.025} \Rightarrow \text{Reject } H_0$$

i) (3pts.) Predict the level of L for the year 2013 when the average rate of GDP growth is 2%.

$$\hat{L}_{2013} = -0.5458 + (0.4857)G_{2013}$$

$$\hat{L}_{2013} = -0.5458 + (0.4857)(2.00) = 0.4336$$

j) (3pts.) Obtain the least squares estimate of the prediction error variance.

$$\sigma_{ep}^2 = \sigma^2 \left(1 + \frac{1}{25} + \left(x_p^2 / \sum x_i^2 \right) \right)$$

$$= 0.4404 \left(1 + \frac{1}{25} + (-0.8168)^2 / 60.7703 \right)$$

$$= 0.4404 (1 + 0.04 + 0.6672) / 60.7703$$

$$= 0.4404 (1 + 0.07 + 0.0110)$$

$$= 0.4628$$

$$\text{where } x_p = 2 - \bar{X} = 2 - 2.8168 = -0.8168$$

k) (3pts.) Construct a 95% confidence interval for L_{2013} when $G_{2013} = 2$.

$$\begin{aligned}\hat{\sigma}_{ep} &= \sqrt{0.4628} = 0.6804, \quad t_{23, 0.025} = 2.068 \\ P[\hat{q}_{2013} - (\hat{\sigma}_{ep})(t_{23, 0.025}) < y_p < \hat{q}_{2013} + (\hat{\sigma}_{ep})(t_{23, 0.025})] &= 0.95 \\ P[0.4336 - (0.6804)(2.068) < y_p < 0.4336 + (0.6804)(2.068)] &= 0.95 \\ P[0.4336 - 1.4077 < y_p < 0.4336 + 1.4077] &= 0.95 \\ P[-0.9741 < y_p < 1.8413] &= 0.95\end{aligned}$$

l) (3pts.) Test whether prediction error is statistically significant when $L_{2013} = 1.5\%$.

$$\begin{aligned}H_0: E(e_p) &= 0 & e_p &= y_p - \hat{q}_p \\ H_1: E(e_p) &\neq 0 & e_p &= 1.5 - 0.4336 = 1.0664 \\ \hat{\sigma}_{ep} &= 0.6804 & t_{23, 0.025} &= 2.068\end{aligned}$$

$$t = 1.0664 / 0.6804 = \boxed{1.5673}$$

$t < t_{23, 0.025} \Rightarrow$ Not reject H_0 .

Prediction error is not statistically significant.

m) (4pts.) Interpret the model you have estimated and explain in detail whether the model should be reformulated.

No need to reformulate model since slope term is both statistically and economically significant. The existing model explains 55% of the relation as shown by R^2 . The model as a whole is significant.

The only problem is related with intercept term. It seems insignificant. The model without intercept term may improve the results.

2. (30 pts.) Consider the following simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, 2, \dots, n$$

where e_i 's are iid $N(0, \sigma^2)$ random variables, x_i 's are known constants (in the sense that they are not random) and β_0 , β_1 and σ^2 are the parameters to be estimated. We know that the OLS estimators for β_0 , β_1 and σ^2 are

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{x}_n, \quad \hat{\beta}_1 = \left[\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right]^{-1} \left[\sum_{i=1}^n (x_i - \bar{x}_n)(Y_i - \bar{Y}_n) \right]$$

and

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

where $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, i = 1, 2, \dots, n$. We also know that all these three estimators are unbiased.

a) (15pts.) Show that the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased for the parameters β_0 and β_1 . Calculate their variances.

Note that

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) (Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2}$$

and

$$E(Y_i) = \beta_0 + \beta_1 x_i, \quad \text{Var}(Y_i) = \sigma^2$$

$$E(\hat{\beta}_1) = \frac{\sum (x_i - \bar{x}) E(Y_i)}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) (\beta_0 + \beta_1 x_i)}{\sum (x_i - \bar{x})^2}$$

$$= \beta_0 \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} + \beta_1 \frac{\sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x})^2} = \beta_1 \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = \beta_1$$

$$\boxed{E(\hat{\beta}_1) = \beta_1}$$

$$\text{Var}(\hat{\beta}_1) = \left[\frac{1}{\sum (x_i - \bar{x})^2} \right]^2 \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$\boxed{\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}}$$

Now

$$\begin{aligned} E(\hat{\beta}_0) &= E(\bar{Y}_n - \hat{\beta}_1 \bar{x}_n) = E(\bar{Y}_n) - \bar{x}_n E(\hat{\beta}_1) \\ &= \beta_0 + \beta_1 \bar{x}_n - \beta_1 \bar{x}_n = \beta_0 \end{aligned}$$

$$\begin{aligned} E(\bar{Y}_n) &= \frac{1}{n} \sum E(Y_i) \\ &= \frac{1}{n} \sum (\beta_0 + \beta_1 x_i) \\ &= \beta_0 + \beta_1 \bar{x}_n \end{aligned}$$

Note that

$$\begin{aligned}\text{Var}(\hat{\beta}_0) &= \text{Var}(\bar{Y}_n - \hat{\beta}_1 \bar{X}_n) \\ &= \text{Var}(\bar{Y}_n) + \bar{X}_n^2 \text{Var}(\hat{\beta}_1) - 2\bar{X}_n \text{Cov}(\bar{Y}_n, \hat{\beta}_1)\end{aligned}$$

However

$$\hat{\beta}_1 = \sum_{i=1}^n c_i Y_i = \sum_{i=1}^n \left(\frac{x_i - \bar{x}_n}{\sum (x_i - \bar{x}_n)^2} \right) Y_i = \sum_{i=1}^n c_i Y_i$$

$$c_i = \frac{x_i - \bar{x}_n}{S_{xx}} \quad \text{and} \quad \sum c_i = 0$$

$$\begin{aligned}\text{Cov}(\bar{Y}_n, \hat{\beta}_1) &= \text{Cov}\left(\frac{1}{n} \sum Y_i, \sum c_i Y_i\right) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n c_i \text{Cov}(Y_i, Y_j) \\ &= \frac{1}{n} \sum_{i=1}^n c_i = 0, \quad Y_i \text{'s are indep.} \\ &\quad \text{Cov}(Y_i, Y_j) = 0 \text{ for } i \neq j\end{aligned}$$

Hence,

$$\begin{aligned}\text{Var}(\hat{\beta}_0) &= \text{Var}(\bar{Y}_n) + \bar{X}_n^2 \text{Var}(\hat{\beta}_1) \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}_n^2}{\sum (x_i - \bar{x}_n)^2} = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}_n^2}{\sum (x_i - \bar{x}_n)^2} \right]\end{aligned}$$

i.e.

$$E(\hat{\beta}_0) = \beta_0$$

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}_n^2}{\sum (x_i - \bar{x}_n)^2} \right]$$

b) (15pts.) Find the maximum likelihood estimator for the parameters β_0 , β_1 and σ^2 . Are these MLE estimators unbiased the β_0 , β_1 and σ^2 ?

if $e_i \sim \text{iid } N(0, \sigma^2)$, then $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

pdf of y_i 's

$$f_{y_i}(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (y_i - \beta_0 - \beta_1 x_i)^2\right\}, y_i \in \mathbb{R}$$

the likelihood function at $\underline{\theta} = (\beta_0, \beta_1, \sigma^2)$ is

$$L(\underline{\theta}, \underline{y}) = \prod_{i=1}^n f(y_i; \underline{\theta}) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

and the log-likelihood function

$$\begin{aligned} \ell(\underline{\theta}) &= \ln(L(\underline{\theta}, \underline{y})) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned}$$

$$\frac{d}{d\beta_0} \ell(\underline{\theta}) = 0 + 0 + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{d}{d\beta_1} \ell(\underline{\theta}) = 0 + 0 + \frac{1}{2\sigma^2} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{d}{d\sigma^2} \ell(\underline{\theta}) = 0 - \frac{n}{2\sigma^2} + \frac{1}{4\sigma^4} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

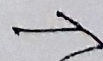
From first two derivatives, we get

$$\left. \begin{aligned} \sum (y_i - \beta_0 - \beta_1 x_i) &= 0 \\ \sum x_i (y_i - \beta_0 - \beta_1 x_i) &= 0 \end{aligned} \right\} \Rightarrow \begin{aligned} n\beta_0 + \beta_1 \sum x_i &= \sum y_i \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 &= \sum x_i y_i \end{aligned}$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

the MLE's are
the same as OLS
(for β_0 & β_1)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$



The MLE estimates for β_0 & β_1 are

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Note that these are the same as OLS estimates and therefore these are unbiased for β_0 & β_1

i.e. $E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

Put these two solutions into the third derivative

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)^2 = 0$$

$$\frac{\hat{\sigma}_{MLE}^2}{n} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Note that

$$\hat{\sigma}_{OLS}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ is unbiased for } \sigma^2$$

$$\therefore E(\hat{\sigma}_{MLE}^2) = E\left(\frac{n-2}{n} \cdot \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2\right)$$

$$= \frac{n-2}{n} E\left(\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2\right)$$

$$= \frac{n-2}{n} \sigma^2$$

$$E(\hat{\sigma}_{MLE}^2) = \frac{n-2}{n} \sigma^2$$

$\hat{\sigma}_{MLE}^2$ is biased for σ^2

3. (30 pts.) Let $X_1, X_2, \dots, X_n$ be a random sample from a population having a probability density function

$$f(x; \theta) = \begin{cases} \theta x^{\theta-1} & , 0 \leq x \leq 1 \\ 0 & . \text{ elsewhere} \end{cases}$$

a) (10 pts.) Find a sufficient estimator for θ (call T)

$$f(\underline{x}; \theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \theta x_i^{\theta-1} = \underbrace{\theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1}}_{g(T(\underline{x}); \theta)} \cdot \underbrace{\left(\prod_{i=1}^n x_i \right)^{-1}}_{h(\underline{x})} = g(T(\underline{x}); \theta) \cdot h(\underline{x})$$

By Factorization thm.

$T = \prod_{i=1}^n x_i$ is sufficient for θ .

Remember that any one-to-one function of a suff. stat. is also suff. so, $T = -\ln \prod_{i=1}^n x_i = -\sum_{i=1}^n \ln x_i$

$T = -\sum_{i=1}^n \ln x_i$ is suff. for θ .

b) (10 pts.) Based on the sufficient estimator (you found in (a)), find an unbiased estimator for θ (call W)

Note that $T = -\sum_{i=1}^n \ln x_i$ is suff. for θ .

let $y_i = -\ln x_i$, $D_x = (0, 1)$, $D_y = \mathbb{R}^+$

let us find the pdf of Y .

Note that $F_Y(y) = 0$ for $y < 0$

for $y > 0$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(-\ln X \leq y) = P(Y > e^{-y}) = \int_{e^{-y}}^1 \theta x^{\theta-1} dx \\ &= x^{\theta} \Big|_{e^{-y}}^1 = 1 - e^{-\theta y} \end{aligned}$$

$$\text{so, } F_Y(y) = \begin{cases} 1 - e^{-\theta y} & , y > 0 \\ 0 & , \text{ else} \end{cases}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \theta e^{-\theta y} & , y > 0 \\ 0 & , \text{ else} \end{cases}$$

$$f_Y(y; \theta) = \begin{cases} \theta e^{-\theta y} & , y > 0 \\ 0 & , \text{ else.} \end{cases}$$

$Y \sim \text{exp}(1/\theta)$



Since

$$Y_i \sim \exp(1/\theta),$$

$$T = -\sum_{i=1}^n \ln X_i = \sum_{i=1}^n Y_i \sim \text{Gamma}(n, 1/\theta)$$

$$f_T(t) = \begin{cases} \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t}, & t > 0 \\ 0, & \text{else.} \end{cases}$$

calculate $E(1/T)$

$$\begin{aligned} E\left(\frac{1}{T}\right) &= \int_0^{\infty} \frac{1}{t} f_T(t) dt = \frac{\theta^n}{\Gamma(n)} \int_0^{\infty} \frac{1}{t} t^{n-1} e^{-\theta t} dt \\ &= \frac{\theta^n}{\Gamma(n)} \int_0^{\infty} t^{n-2} e^{-\theta t} dt = \frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-1)}{\theta^{n-1}} \\ &= \theta \frac{\Gamma(n-1)}{\Gamma(n)} = \theta \frac{\Gamma(n-1)}{(n-1)\Gamma(n-1)} = \frac{\theta}{n-1} \end{aligned}$$

Since

$$E\left(\frac{1}{T}\right) = \frac{\theta}{n-1}$$

$W = \frac{n-1}{T}$ is unbiased for θ .

i.e.

$$E(W) = E\left(\frac{n-1}{T}\right) = (n-1) E\left(\frac{1}{T}\right) = (n-1) \frac{\theta}{n-1} = \theta$$

$W = \frac{n-1}{T}$ is unbiased for θ .

c) (10 pts.) Is the unbiased estimator you found in (b) UMVUE for θ ?

Note that T is suff. for θ and since

$T \sim \text{Gamma}(n, 1/\theta)$, T is complete.

W is unbiased for θ .

By Rao-Blackwell theorem

$\phi(T) = E(W|T)$ is the unique UMVUE for θ

$$\phi(T) = E(W|T) = E\left(\frac{n-1}{T} | T\right) = \frac{n-1}{T} = W$$

i.e.

$$\boxed{W = \frac{n-1}{T}}$$

$\Rightarrow$ the UMVUE for θ .

here

$$T = - \sum_{i=1}^n \ln X_i$$

NAME:.....

SPRING 2020-2021

ECON 206

ERDIL, AKDI

08.07.2021

FINAL EXAM
(TIME ALLOWED 130 MIN)
NO PARTIAL CREDIT

1. (50pts.) Consider the following model for the relation between median home prices and 10-year mortgage interest rates in Turkey for the period 1988-2012.

$$P_t = \beta_0 + \beta_1 R_t + u_t \quad t = 1988, \dots, 2012.$$

where P denotes median home prices and R is 10-year mortgage interest rates.

You are provided the following information from the sample:

$$\sum_{t=1988}^{2012} P_t = 6,081.10 \quad \sum_{t=1988}^{2012} R_t = 181.85 \quad \sum_{t=1988}^{2012} P_t R_t = 42,193.12$$

$$\sum_{t=1988}^{2012} P_t^2 = 1,593,606.07 \quad \sum_{t=1988}^{2012} R_t^2 = 1,377.89 \quad \sum_{t=1988}^{2012} p_t r_t = -2,040.80$$

$$\sum_{t=1988}^{2012} p_t^2 = 114,414.98 \quad \sum_{t=1988}^{2012} r_t^2 = 55.11$$

$$\text{where } p_t = (P_t - \bar{P}) \quad r_t = (R_t - \bar{R})$$

LET $\alpha = 0.05$ FOR ALL TESTS OF SIGNIFICANCE UNLESS OTHERWISE STATED.

a) (4 pts.) Obtain the least squares estimates of β_0 and β_1 .

$$\sum (P_t - \bar{P})(R_t - \bar{R}) = -2040.80$$

$$\sum (R_t - \bar{R})^2 = 55.11$$

$$\bar{P} = \frac{6081.10}{25} = 243.24$$

$$\bar{R} = \frac{181.85}{25} = 7.274$$

b) (4 pts.) Calculate the coefficient of determination (R^2), simple correlation coefficient between P_t and $\widehat{P}_t$ ($r_{\widehat{P}_t}$), and simple correlation coefficient between P_t and R_t (r_{R_t}) then interpret them.

c) (4 pts.) Obtain the least squares estimate of the variance of the model.

d) (4 pts.) Obtain the least squares estimates of the variances of the coefficients.

e) (4 pts.) Construct a 95% confidence interval for β_0 .

f) (4 pts.) Evaluate whether slope coefficient is both statistically and economically significant.

g) (4 pts.) Test whether the regression model as a whole is significant.

h) (4 pts.) Test if $\beta_0 + \beta_1 = 1$.

i) (4 pts.) Predict the level of P for the year 2013 when the average rate of interest is 7%.

j) (4 pts.) Obtain the least squares estimate of the prediction error variance.

k) (3 pts.) Construct a 95% confidence interval for P_{2013} when $R_{2013} = 7$.

l) (4 pts.) Test whether prediction error is statistically significant when $P_{2013} = 260$.

m) (3 pts.) Interpret the model you have estimated and explain in detail whether the model should be reformulated.

2. (10 pts.)

a) (8 pts.) Complete the following ANOVA table for a simple linear regression of Y on X (that is calculate the values A,B,C and D).

SoV	df	SS	MS	F
Regression	1	A	0.178225	58.956
Error	B	0.0090703	D	
Total	C	0.1872928		

b) (2 pts.) Calculate R^2 .

3. (25 pts.) Assume that the daily number of breads (X) sold in a grocery store is normally distributed with mean μ and variance σ^2 . Regarding the 25-day bread sales we observe that

$$\sum_{i=1}^{25} x_i = 3000 \text{ and } \sum_{i=1}^{25} x_i^2 = 363456.$$

- a) (10 Points) Test the null hypothesis that the number of daily sales is less than 115 breads against the alternative that the number of breads sold is more than 115 at the 5% significance level.

- b) (5 Points) Now, let's assume that $\sigma^2 = 144$ and test the same hypothesis given in (a).

- c) **(10 Points)** According to the hypothesis testing problem given in (b) calculate the power of the test for $\mu = 110, 115$ and 120

4. (15 Points) Consider the simple linear regression model given below:

$$Y_i = \alpha + \beta x_i + e_i, \quad i = 1, 2, \dots, n.$$

Show that

$$\sum_{i=1}^n \hat{e}_i = 0$$

$$\sum_{i=1}^n \hat{Y}_i \hat{e}_i = 0$$

$$\text{and} \quad \sum_{i=1}^n x_i \hat{e}_i = 0$$

where $\hat{Y}_i$ and $\hat{e}_i$ denote the predicted values and residuals respectively.