Name: ....
Dept: ....

Date: May 22, 2013
Time: 15:30-17:30
Instructor: Dilek Güvenç

MATH 260 FINAL EXAM

1	2	3	4	5	TOTAL
15	25	20	20	20	100

- Check that there are 5 questions in your booklet.
- Show all your work. Correct results without sufficient explanation and correct notation may not get full credit. Please, be neat.
- Write your name on each page of your booklet.
- Do NOT use your mobile phone as a calculator, turn it off during the exam.
- Do NOT forget to return the tables and the "formula page", as well.

********* **GOOD LUCK & HAVE A NICE SUMMER!** *********

---

1. Let X_i's ($i = 1,2, \dots n$) denote independent observed values from a normal distribution with mean μ_i, each and variance σ^2, and Y_i values denote independent observed values from the same distribution (normal with mean μ_i, variance σ^2). Suppose X_i and Y_i are independent of one another.

a) Use $Z_i = X_i - Y_i$ to find maximum likelihood estimator, $\hat{\sigma}^2$ of σ^2. **(10 Points)**

b) Is maximum likelihood estimator, $\hat{\sigma}^2$ an unbiased estimator of σ^2? Why? **(5 Points)**

a) $X_i \sim N(\mu_i, \sigma^2)$, $Y_i \sim N(\mu_i, \sigma^2) \Rightarrow X_i - Y_i \cancel{is}\ Z_i$'s normal.

$E(Z_i) = E(X_i - Y_i) = E(X_i) - E(Y_i) = \mu_i - \mu_i = 0$

$\sigma^2_{Z_i} = Var(Z_i) = Var(X_i - Y_i) = Var(X_i) + Var(Y_i) = \sigma^2 + \sigma^2 = 2\sigma^2$

$\Rightarrow Z_i \sim N(0, 2\sigma^2)$

$$f(z, \sigma^2) = \frac{1}{\sqrt{2\sigma^2}\sqrt{2\pi}} e^{-\frac{1}{2}\frac{z^2}{2\sigma^2}} = \frac{1}{\sqrt{\sigma^2}\, 2\sqrt{\pi}} e^{-\frac{z^2}{4\sigma^2}}$$

$$L_z(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma^2}\, 2\sqrt{\pi}} e^{-\frac{z_i^2}{4\sigma^2}} \Rightarrow K_z(\sigma^2) = \sum_{i=1}^{n}\left(-\ln 2\sqrt{\pi} - \frac{1}{2}\ln\sigma^2 - \frac{z_i^2}{4\sigma^2}\right)$$

$$= -n\ln 2\sqrt{\pi} - \frac{n}{2}\ln\sigma^2 - \frac{\sum z_i^2}{4\sigma^2}$$

$$\frac{dK_z(\sigma^2)}{d\sigma^2} = 0 \Rightarrow -\frac{n}{2\sigma^2} + \frac{\sum z_i^2}{4(\sigma^2)^2} = 0 \Rightarrow -2n\sigma^2 + \sum z_i^2 = 0$$

$\Rightarrow$ M.L.E of σ^2,
$$\hat{\sigma}^2 = \frac{\sum z_i^2}{2n} = \frac{\sum (X_i - Y_i)^2}{2n}$$

$\longrightarrow$

b) $E(\hat{\sigma}^2) = E\left(\frac{\sum z_i^2}{2n}\right)$

$$= \frac{1}{2n} \sum_{i=1}^{n} E(z_i^2)$$

$$= \frac{1}{2n} \sum_{i=1}^{n} \left[Var(z_i) + (E(z_i))^2 \right]$$

$$= \frac{1}{2n} \sum_{i=1}^{n} (2\sigma^2 + 0)$$

$$= \frac{1}{\cancel{2n}} \cancel{n}\cancel{2}\sigma^2 = \sigma^2$$

$\Rightarrow \hat{\sigma}^2$ is an unbiased estimator of σ^2.

——— o ———

Name:................

2. The concentration of an active ingredient in the output of chemical reaction is strongly influenced by the catalyst that is used in the reaction. It is felt that when catalyst A is used, the population mean concentration is at least 0.65. The standard deviation (σ) is known to be 0.05. A sample of outputs from 31 independent experiments gives the average concentration of 0.645.

a) Does this sample provide sufficient information to suggest that true mean concentration for catalyst A is less than 0.65? Test at $\alpha = 0.01$ and use P-value to make the decision. **(8 Points)**

b) Suppose a similar experiment is done with the use of another catalyst, catalyst B, standard deviation is still assumed to be 0.05 and 31 independent experiments with catalyst B gives the average concentration of 0.70. Can we conclude that true mean difference between catalyst B and catalyst A concentrations is more than 5%? Test at $\alpha = 0.05$. **(8 Points)**

c) Find a 90% confidence interval for the true mean difference between catalystAand catalyst B concentrations. **(3 Points)**

d) Suppose that concentration levels for both type of catalyst are normally distributed, for the experiments described above, find

$$P(\frac{S_A}{S_B} > 1.27).$$

(6 Points)

a) $\bar{X}_A = 0.645 \quad \sigma = 0.05 \quad n_A = 31$

$H_0: \mu_A \geq 0.65$ Under H_0, $Z_{cal} = \dfrac{0.645 - 0.65}{\frac{0.05}{\sqrt{31}}} = -0.55$

$H_A: \mu_A < 0.65$

P-Value $= P(Z < Z_{cal}) = P(Z < -0.55) = 0.5 - 0.2088 = 0.2912$

Since P-Value $= 0.2912 > 0.01$ we do not reject H_0, conclude that μ_A might be at least 0.65.

b) $\bar{X}_B = 0.70 \quad n_B = 31 \quad \sigma = 0.05$.

$H_0: \mu_B - \mu_A = 0.05$ $\quad Z_{cal} = \dfrac{0.70 - 0.645 - 0.05}{0.05\sqrt{\frac{1}{31} + \frac{1}{31}}} = \dfrac{0.005}{0.0127} = 0.39$

$H_A: \mu_B - \mu_A > 0.05$

$Z_{0.05} = 1.645$ Since $Z_{cal} = 0.39 \not> Z_{0.05} = 1.645$ we do not reject H_0, conclude that difference may not be > 0.05 at $\alpha = 0.05$.

c) 90% C.I for $\mu_B - \mu_A$

$0.70 - 0.645 \pm (1.645)(0.05)\sqrt{\frac{2}{31}}$ or 0.055 ± 0.021

d) $P(\frac{S_A}{S_B} > 1.27) = P(\frac{S_A^2}{S_B^2} > 1.61) = P(F(30,30) > 1.61) \sim 0.10$

Because $S_A^2 / S_B^2 = \dfrac{30 S_A^2 / 30 (0.05)^2}{30 S_B^2 / 30 (0.05)^2} \sim F(30,30)$

Name:..................

3. In simple regression show that

a) $E(\overline{Y}) = \beta_0 + \beta_1 \overline{x}$ **(6 Points)**

b) $\hat{\beta}_1$ is an unbiased estimator of β_1. **(10 Points)**

c) $\hat{\beta}_0$ is an unbiased estimator of β_0. **(4 Points)**

$$a)\ E(\overline{Y}) = E\left(\frac{\sum Y_i}{n}\right) = \frac{1}{n}\sum E(Y_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 X_i)$$
$$= \frac{1}{n}(n\beta_0 + \beta_1 \sum X_i)$$
$$= \beta_0 + \beta_1 \frac{\sum X_i}{n}$$
$$= \beta_0 + \beta_1 \overline{X}$$

$$b)\ \hat{\beta}_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum (X_i - \overline{X})Y_i - \overline{Y}\overbrace{\sum(X_i - \overline{X})}^{0}}{\sum (X_i - \overline{X})^2}$$

$$= \frac{\sum (X_i - \overline{X})}{\sum (X_i - \overline{X})^2} Y_i$$

$$E(\hat{\beta}_1) = \frac{\sum (X_i - \overline{X})}{\sum (X_i - \overline{X})^2} E(Y_i) = \frac{\sum (X_i - \overline{X})(\beta_0 + \beta_1 X_i)}{\sum (X_i - \overline{X})^2}$$
$$= \frac{\beta_0 \overbrace{\sum (X_i - \overline{X})}^{0} + \beta_1 \sum (X_i - \overline{X})X_i}{\sum (X_i - \overline{X})^2}$$
$$= \frac{\beta_1 [\sum X_i^2 - \overline{X}\sum X_i]}{\sum (X_i - \overline{X})^2} = \frac{\beta_1 (\cancel{\sum X_i^2 - n\overline{X}^2})}{\cancel{\sum X_i^2 - n\overline{X}^2}} = \beta_1$$

$\Rightarrow \hat{\beta}_1$ is an unbiased estimator of β_1

$$c)\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$
$$E(\hat{\beta}_0) = E(\overline{Y} - \hat{\beta}_1 \overline{X}) = E(\overline{Y}) - \overline{X} E(\hat{\beta}_1)$$
$$= \beta_0 + \cancel{\beta_1 \overline{X}} - \cancel{\beta_1 \overline{X}}$$
$$= \beta_0$$

$\Rightarrow \hat{\beta}_0$ is an unbiased estimator of β_0

Name:.....................

4. According to a part of computer output in the table below,

a) Predict the single response when $x = 5000$. **(2 Points)**

b) Is x significant variable for the model? Why? **(3 Points)**

c) Estimate the error standard deviation. **(3 Points)**

You will need the additional information to answer parts (c) and (d) :

$\bar{x} = 8354$ and $S_{xx} = 97\,599\,296$.

d) Estimate the mean response when $x = 5000$ with a 90% confidence interval. **(3 Points)**

e) Predict single response with a 90 % interval when $x = 5000$. **(3 Points)**

f) Find 95 % confidence interval for the slope of true regression line. **(3 Points)**

g) What is the proportion of Y-variability explained by this model? **(3 points)**

THE REGRESSION EQUATION IS

Y = 994 + 0.104 x

PREDICTOR	COEF	STDEV	T-RATIO	P
CONSTANT	994	254.7	3.9	0.001
X	0.10373	0.02978	3.48	0.002

ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS	F	P
REGRESSION	1	1087765	1087765	12.14	0.002
ERROR	28	2509820	89636		
TOTAL	29	3597585			

a) $\hat{y} = 994 + (0{,}104)5000 = 994 + 520 = 1514$

b) Yes, P-Value for the test $H_0: \beta_1 = 0$ versus $H_0: \beta_1 \neq 0$ is $0{,}002 < 0{,}05$.

c) $s^2 = MSE = 89636 \Rightarrow s = 299.39$

d) Since $n-1 = 29$, $n = 30$ $\quad t_{0,05}(28) = 1{,}701$

90 % C.I for $E(Y)$

$994 \pm (0{,}14)5000 \pm (1{,}701)(299{,}39)\sqrt{\frac{1}{30} + \frac{11249316}{97599296}}$

$1514 \pm 196{,}3$ or $(1317{,}69,\ 1710{,}3)$

e) 90 % prediction interval if $x = 5000$

$1514 \pm (1{,}701)(299{,}39)\sqrt{1 + \frac{1}{30} + \frac{11249316}{97599296}}$

$1514 \pm 545{,}79$ or $(968{,}21,\ 2059{,}79)$

f) 95 % C.I for β_1 $\quad 0{,}1037 \pm t_{0{,}025}(28)(0{,}02978)$

$0{,}1037 \pm (2{,}048)(0{,}02978)$

g) $R^2 = \frac{SSR}{SST} = \frac{1087765}{3597585} = 0{,}3023 \quad 30{,}2\%$

Name:................

5. The octane number of refined petroleum is related to the temperature x of the refining process, but it is also related to the particle size of catalyst. An experiment with a small-particle catalyst gave a fitted regression line of

$$\hat{y} = 9.36 + 0.115x$$

with n = 21 and $s_{\hat{\beta}_1} = 0.0225$. An independent experiment with a large-particle catalyst gave

$$\hat{y} = 4.265 + 0.19x$$

with n = 11 and $s_{\hat{\beta}_1} = 0.0202$.

a) For small-particle catalyst, test the hypothesis that the slope is significantly different from 0 at 5% significance level.**(5 Points)**

b) Determine the test statistic for the hypothesis that two types of catalyst produce the same slope in the relationship and test at 5% level of significance. Temperatures of refining processes were the same for both types of catalysts and assume that true error variances are equal.
(15 Points)

a) $H_0: \beta_1 = 0$ Under H_0,

$H_A: \beta_1 \neq 0$ $T_{cal} = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = \frac{0.115}{0.0225} = 5.11$

$t_{0.025}^{(19)} = 2.093$ Since $T_{cal} = 5.11 > t_{0.025}^{(19)} = 2.093$ we reject H_0. Conclude that $\beta_1 \neq 0$ at $\alpha = 0.05$

b) $H_0: \beta_1 - \beta_2 = 0$ β_1: slope of small-particle model

$H_A: \beta_1 - \beta_2 \neq 0$ β_2: " " large-particle "

Estimator of $\beta_1 - \beta_2$ is $\hat{\beta}_1 - \hat{\beta}_2$.

$Var(\hat{\beta}_1 - \hat{\beta}_2) = \frac{\sigma^2}{Sxx} + \frac{\sigma^2}{Sxx} = \frac{2\sigma^2}{Sxx}$ temperatures (x's) are the same for both models

σ^2 is unknown common variance, we need to find its estimator using the data obtained from both experiments

$$S^2_{pooled} = \frac{SSE_1 + SSE_2}{n_1 + n_2 - 4} = \frac{(n_1 - 2)S_1^2 + (n_2 - 2)S_2^2}{n_1 + n_2 - 4}$$

In the question $S_{\hat{\beta}_1}$ and $S_{\hat{\beta}_2}$ are given we write S^2_{pooled} in terms of these $S_{\hat{\beta}_1} = \frac{S_1}{\sqrt{Sxx}} \Rightarrow S_1^2 = S^2_{\hat{\beta}_1} Sxx$

→

Therefore estimator of $Var(\hat{\beta}_1 - \hat{\beta}_2)$

$$S^2_{\hat{\beta}_1 - \hat{\beta}_2} = \frac{2S^2_{pooled}}{S_{XX}}$$

$$= 2\left[\frac{(n_1-2)S^2_{\hat{\beta}_1}S_{XX} + (n_2-2)S^2_{\hat{\beta}_2}S_{XX}}{n_1+n_2-4}\right]\frac{1}{S_{XX}}$$

$$= 2\,\frac{(n_1-2)S^2_{\hat{\beta}_1} + (n_2-2)S^2_{\hat{\beta}_2}}{n_1+n_2-4}$$

$$= \frac{2\,[19(0.0225)^2 + 9(0.0202)^2]}{21+11-4}$$

$$= 0.000949 \implies S_{\hat{\beta}_1-\hat{\beta}_2} = 0.0308$$

$$\implies T_{cal} = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 0}{S_{\hat{\beta}_1-\hat{\beta}_2}} = \frac{0.115 - 0.19}{0.0308} = -2.43$$

$d.o.f = 21+11-4 = 28 \qquad t_{0.025}(28) = 2.048$

Since $|T_{cal}| = 2.43 > t_{0.025}(28) = 2.048$ we reject H_0, conclude that the two types of catalysts give regression lines with significantly different slopes at $\alpha = 0.05$.