AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

1. A research team conducted a pilot study of whether a training program improved the productivity of a firm's workers. They took a random sample of $n = 5$ workers and measured their productivity before and after the training program. The average productivity increase was $\bar{y}_5 = 15.5$ units, and the sample variance of the productivity increase was 98.3 (the divisor in the variance was $n-1$). Test the null hypothesis that $E(Y) = 0$ against the alternative that $E(Y) \neq 0$. Use levels of significance 0.10, 0.05, and 0.01. This problem is worth 40 points.

$$t_4 = \frac{\bar{y}_5 - 0}{\sqrt{\frac{\sigma_4^2}{5}}} = \frac{15.5 - 0.}{\sqrt{\frac{98.3}{5}}} = \frac{15.5}{\sqrt{19.66}} = \frac{15.5}{4.434} = 3.496$$

α	t_4	
.10	2.132	REJECT
.05	2.776	REJECT
.01	4.604	ACCEPT.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

2. A research team took a random sample of 3 observations from a normally distributed random variable Y and observed that $\bar{y}_3 = 44.9$ and $s_Y^2 = 239.7$, where $\bar{y}_3$ was the average of the three observations sampled from Y and s_Y^2 was the unbiased estimate of $\text{var}(Y)$ (i.e., the divisor in the variance was $n-1$). A second research team took a random sample of 6 observations from a normally distributed random variable X and observed that $\bar{x}_6 = 81.2$ and $s_X^2 = 235.1$, where $\bar{x}_6$ was the average of the six observations sampled from X and s_X^2 was the unbiased estimate of $\text{var}(X)$ (i.e., the divisor in the variance was $n-1$). Calculate the 99% confidence interval for $E(X) - E(Y)$ using the pooled variance estimator. This problem is worth 40 points.

$$S_p^2 = \frac{2(239.7) + 5(235.1)}{7} = \frac{1654.9}{7} = 236.414$$

99% CI FOR $E(X) - E(Y)$

$$81.2 - 44.9 \mp 3.499\sqrt{(236.414)\left(\tfrac{1}{3} + \tfrac{1}{6}\right)}$$

$$36.3 \pm 3.499\sqrt{118.207}$$

$$36.3 \pm 38.04$$

$$-1.74 \text{ TO } 74.34$$

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

3. A research team took a random sample of 5 observations from a normally distributed random variable Y and observed that $\bar{y}_5 = 831.8$ and $s_Y^2 = 910.6$, where $\bar{y}_5$ was the average of the five observations sampled from Y and s_Y^2 was the unbiased estimate of $\text{var}(Y)$ (i.e., the divisor in the variance was $n-1$). A second research team took a random sample of 6 observations from a normally distributed random variable X and observed that $\bar{x}_6 = 172.5$ and $s_X^2 = 325.2$, where $\bar{x}_6$ was the average of the six observations sampled from X and s_X^2 was the unbiased estimate of $\text{var}(X)$ (i.e., the divisor in the variance was $n-1$). Test the null hypothesis $H_0 : \text{var}(Y) = \text{var}(X)$ against the alternative $H_1 : \text{var}(Y) > \text{var}(X)$ at the 0.10, 0.05, and 0.01 levels of significance. This problem is worth 40 points.

$$TS = \frac{VAR(Y)}{VAR(X)} = \frac{910.6}{325.2} = 2.80 \text{ ON } (4,5) \text{DF}$$

α	$F(4,5)$	
.10	3.52	ACCEPT
.05	5.19	ACCEPT
.01	11.39	ACCEPT

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

4. A research team collected data on $n = 103$ participants in a longitudinal study. Each participant reported average weekly alcohol consumption at age 22. The average alcohol consumption reported at age 22 was 4.2 units, with an observed standard deviation of 2.36 units (the divisor in the underlying variance calculation was $n-1$). Each participant subsequently completed a questionnaire on depression symptoms at age 25. The reported average number of depression symptoms at age 25 was 6.12, with an observed standard deviation of 3.87 (the divisor in the underlying variance calculation was $n-1$). The Pearson product moment correlation coefficient between the two variables was 0.17. The research team seeks to estimate the regression of number of depression symptoms at age 25 on alcohol consumption at age 22.

a. Find the estimated regression equation of number of depression symptoms at age 25 on alcohol consumption at age 22. Find the 99% confidence interval for the slope in this equation. (15 points).
b. Complete the analysis of variance table for this regression and test the null hypothesis that the slope is zero at levels of significance 0.10, 0.05, and 0.01. (15 points)
c. Use the least-squares prediction equation to estimate the number of depression symptoms at age 25 for participants whose alcohol consumption at age 22 was 6.0 units. Give the 99% confidence interval for the number of depression symptoms at age 25 for the participants whose alcohol consumption at age 22 was 6.0 units. (20 points)

B. $TSS = 102(SD(DV))^2 = 102(3.87)^2 = 1527.64$

SOURCE	DF	SS	MS	F
REG	1	44.149	44.149	3.00
ERROR	101	1483.491	14.688	
TOTAL	102	1527.64		

α	$F(1,101)$	
.10	2.755	REJECT
.05	3.933	ACCEPT
.01	6.888	ACCEPT

A $\hat{\beta}_1 = (0.17)\left(\frac{3.87}{2.36}\right) = 0.279$

$\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$
$= 6.12 - 0.279(4.2)$
$= 4.9482$

$\Sigma(x_i - \bar{x}_n)^2 = S_{xx} = 102(2.36)^2$
$= 568.0992$

99% CI FOR β_1: $0.279 \pm 2.625\sqrt{\frac{14.688}{568.0992}}$

$0.279 \pm 2.625\sqrt{.02585} = 0.279 \pm 0.422$

-0.143 TO 0.701

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

C $\hat{Y}_i(6.0) = 4.9482 + 0.279(6) = 6.6222$

99% CI FOR $\beta_0 + 6\beta_1$:

$6.6222 \pm 2.625 \sqrt{14.688\left(\frac{1}{103} + \frac{(6.0-4.2)^2}{568.0992}\right)}$

$6.6222 \pm 2.625 \sqrt{14.688(.009709 + .005703)}$

$6.6222 \pm 2.625 \sqrt{0.2264}$

6.6222 ± 1.249

5.37 TO 7.87

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

5. In a clinical trial, $2J$ patients suffering from an illness will be randomly assigned to one of two groups so that J will receive an experimental treatment and J will receive the best available treatment. The random variable X is the response of a patient to the experimental medicine, and the random variable B is the response of a patient to the best currently available treatment. Both X and B are normally distributed with $\sigma_X = \sigma_B = 500$. The null hypothesis to be tested is that $E(X) - E(B) = 0$ against the alternative that $E(X) - E(B) > 0$ at the 0.005 level of significance. What is the number J in each group that would have to be taken so that the probability of a Type II error for the test of the null hypothesis specified in the common section is 0.01 when $E(X) - E(B) = 400$ and $\sigma_X = \sigma_B = 500$? What is the total number of subjects for this clinical trial? This problem is worth 40 points.

$$\sqrt{J} \geq \frac{2.576\sqrt{2}\,500 + 2.326\sqrt{2}\,500}{400}$$

$$= \frac{1821.507 + 1644.730}{400}$$

$$\sqrt{J} \geq 8.665$$

$$J \geq 76$$

TOTAL # IS $2J \geq 151$.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

6. A research team has taken a random sample $y_1, y_2, \ldots, y_n$ of n observations from a random variable that is normally distributed with expected value μ and variance σ^2 (that is, from a $N(\mu, \sigma^2)$ random variable). They wish to know the ordinary least squares estimate of μ.
   a. Specify the function that is to be optimized to find the estimate (10 points).
   b. Find the ordinary least squares estimate of μ. Prove your estimate (30 points).

End of Examination

a $SS(m) = \sum_{i=1}^{n} (y_i - m)^2.$

b. $\frac{dSS}{dm} = \sum -2(y_i - m).$

NORMAL EQN

$\sum -2(y_i - \hat{\mu}) = 0$

$\hat{\mu} = \bar{y}_n.$

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

1. A research team conducted a pilot study of whether a training program improved the productivity of a firm's workers. They took a random sample of $n = 2$ workers and measured their productivity before and after the training program. The average productivity increase was $\bar{y}_2 = 11.3$ units, and the variance of the productivity increase was 24.8 (the divisor in the variance was $n-1$). What is the 99 % confidence interval for the expected increase? This problem is worth 40 points.

$$\bar{y}_2 \pm t_{2.576,1}\sqrt{\frac{24.8}{2}}$$

$$11.3 \pm 63.657\sqrt{12.4}$$

$$11.3 \pm 63.657(3.52)$$

$$11.3 \pm 224.16 = -212.85 \text{ TO } 235.46$$

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

2. A research team took a random sample of 6 observations from a normally distributed random variable Y and observed that $\bar{y}_6 = 301.8$ and $s_Y^2 = 145.7$, where $\bar{y}_6$ was the average of the six observations sampled from Y and s_Y^2 was the unbiased estimate of $\text{var}(Y)$ (i.e., the divisor in the variance was $n-1$). A second research team took a random sample of 7 observations from a normally distributed random variable X and observed that $\bar{x}_7 = 320.6$ and $s_X^2 = 168.1$, where $\bar{x}_7$ was the average of the seven observations sampled from X and s_X^2 was the unbiased estimate of $\text{var}(X)$ (i.e., the divisor in the variance was $n-1$). Test the null hypothesis $H_0 : E(X) = E(Y)$ against the alternative $H_1 : E(X) \neq E(Y)$ at the 0.10, 0.05, and 0.01 levels of significance using the pooled variance t-test. This problem is worth 40 points.

$$s_p^2 = \frac{5(145.7) + 6(168.1)}{11} = \frac{1737.1}{11} = 157.918$$

$$t_{11} = \frac{320.6 - 301.8}{\sqrt{157.918\left(\frac{1}{7} + \frac{1}{6}\right)}} = \frac{18.8}{\sqrt{157.918(.3095)}}$$

$$= \frac{18.8}{\sqrt{48.879}} = \frac{18.8}{6.991} = 2.69$$

α	Z	t_{11}	
.10	1.645	1.796	R
.05	1.960	2.201	R
.01	2.576	3.106	A

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

3. A research team took a sample of 3 observations from the random variable Y, which had a normal distribution $N(\mu,\sigma^2)$. They observed $\bar{y}_3 = 67.1$, where $\bar{y}_3$ was the average of the three sampled observations and $s^2 = 393.7$ was the observed value of the unbiased estimate of σ^2 based on the sample values (i.e., the divisor in the variance was $n-1$). Find the 99% confidence interval for σ^2. This problem is worth 40 points.

DF = 2

$P(.01003 < \chi^2_2 < 10.60) = 0.99$.

$P\left(0.01003 < \frac{\sum(Y_i - \bar{Y}_3)^2}{\sigma^2} < 10.60\right) = 0.99$

$P\left(\frac{1}{10.60} < \frac{\sigma^2}{2S^2} < \frac{1}{0.01003}\right) = 0.99$

$P\left(\frac{2S^2}{10.60} < \sigma^2 < \frac{2S^2}{0.01003}\right) = 0.99$.

99% CI FOR σ^2: 74.28 TO 78504.49.

IV DV

$t_{2.576, 633} = 2.584$.

	IV	DV
MEAN	3.8	6.12
SD	1.96	4.27

$= 2435.5744$

$\Sigma(x_i - \bar{x}_n)^2 = 634(1.96)^2$

$TSS = 634(4.27)^2$

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

4. A research team collected data on $n = 635$ participants in a longitudinal study. Each participant reported average alcohol consumption per week at age 22. The average alcohol consumption reported at age 22 was 3.8 units, with an observed standard deviation of 1.96 units (the divisor in the underlying variance calculation was $n-1$). Each participant subsequently completed a questionnaire on depression symptoms at age 25. The reported average number of depression symptoms at age 25 was 6.12, with an observed standard deviation of 4.27 (the divisor in the underlying variance calculation was $n-1$). The Pearson product moment correlation coefficient between the two variables was 0.15. The research team seeks to estimate the regression of number of depression symptoms at age 25 on alcohol consumption at age 22.
   a. Find the estimated regression equation of number of depression symptoms at age 25 on alcohol consumption at age 22. Find the 99% confidence interval for the slope in this equation. (15 points).
   b. Complete the analysis of variance table for this regression and test the null hypothesis that the slope is zero at levels of significance 0.10, 0.05, and 0.01. (15 points)
   c. Use the least-squares prediction equation to estimate the number of depression symptoms at age 25 for a participant whose alcohol consumption at age 22 was 7.0 units. Give the 99% prediction interval for the number of depression symptoms at age 25 for this participant (whose alcohol consumption at age 22 was 7.0 units). (20 points)

B

SOURCE	DF	SS	MS	F
REG	1	260.092	260.092	14.57
ERROR	633	11,299.566	17.851	
TOTAL	634	11559.6586		

α	$F(1, 633)$	
.10	2.713	R
.05	3.856	R
.01	6.675	R

A. $\hat{\beta}_1 = (0.15)\dfrac{4.27}{1.96} = 0.3268$, $\hat{\beta}_0 = 6.12 - 0.3268(3.8) = 4.878$

$\hat{Y}(x) = 4.878 + 0.3268\,x.$

99% CI FOR β_1: $0.3268 \pm 2.584\sqrt{\dfrac{17.851}{2435.5744}}$

$0.3268 \pm 2.584\sqrt{.007329} = 0.3268 \pm 0.2212$

$= .1056$ TO 0.548.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

4C). $\hat{y}(7) = 4.878 + 0.3268(7) = 4.878 + 2.2876 = 7.1656$

$$PME = 2.584\sqrt{17.851}\sqrt{1 + \frac{1}{635} + \frac{(7-3.8)^2}{2435.5744}}$$

$$= (2.584)(4.225)\sqrt{1 + 0.001575 + 0.004204}$$

$$= (2.584)(4.225)\sqrt{1.005779}$$

$$= (2.584)(4.225)(1.00288).$$

$$= 10.95.$$

99% PI FOR $Y_F(7)$: 7.1656 ± 10.95

(-3.78) TO 18.11

TRUNCATE TO 0. OK.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

5. A research team wishes to test the null hypothesis $H_0: \rho = 0$ at $\alpha = 0.005$ against the alternative $H_1: \rho > 0$ using Fisher's transformation of the Pearson product moment correlation coefficient as the test statistic. They have asked their consulting statistician for a sample size n such that $\beta = 0.01$ when $\rho = 0.15$ (that is, $\rho^2 = 0.0225$). What is this value? (This question is worth 40 points).

$F(0.15) = \frac{1}{2} \ln\left(\frac{1.15}{0.85}\right) = \frac{1}{2} \ln(1.3529) = \frac{.30228}{2}$

$= 0.1511$

$\sqrt{n-3} \geq \frac{2.576 + 2.326}{0.1511} = 32.43$

$n - 3 \geq 1052$

$n \geq 1055.$

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

6. A research team has taken a random sample $y_1, y_2, \ldots, y_n$ of n observations from a random variable that is normally distributed with expected value μ and variance σ^2 (that is, from a $N(\mu, \sigma^2)$ random variable). They wish to know the ordinary least squares estimate of μ.
    a. Specify the function that is to be optimized to find the estimate (10 points).
    b. Find the ordinary least squares estimate of μ. Prove your estimate (30 points).

End of Examination

a) $SS(m) = \sum_{i=1}^{n} (y_i - m)^2$

b. $\frac{\partial SS}{\partial m} = \sum -2(y_i - m).$

NORMAL EQN

$\sum -2(y_i - \hat{\mu}) = 0.$

$\hat{\mu} = \bar{y}_n.$