Master’s Written Examination
Option: Statistics and Probability

Full points may be obtained for correct answers to eight questions. Each numbered question (which may have several parts) is worth the same number of points. All answers will be graded, but the score for the examination will be the sum of the scores of your best eight solutions.

Use separate answer sheets for each question. DO NOT PUT YOUR NAME ON YOUR ANSWER SHEETS. When you have finished, insert all your answer sheets into the envelope provided, then seal it.
Problem 1—Stat 401. Let \((A_i)_{i=1}^n\) be a sequence of events such that

(i) At least one of these events happens;

(ii) No more than two of these events would happen at the same time.

Suppose \(P(A_i) = p\) for all \(1 \leq i \leq n\), and \(P(A_i \cap A_j) = q\) for all \(i \neq j\). Prove: \(p \geq \frac{1}{n}, q \leq \frac{2}{n}\).

Solution to Problem 1. By (i), \(1 = P(\Omega) = P(\bigcup_{i=1}^n A_i) \leq np\), therefore \(p \geq 1/n\). For the second claim, we prove it by contradiction. We use “\(\sqcup\)” to denote disjoint unions. Assume \(q > 2/n\).

Case 1. \(n\) is even. Then \(1 \geq P[(A_1 \cap A_2) \sqcup (A_3 \cap A_4) \sqcup \cdots \sqcup (A_{n-1} \cap A_n)] \overset{(ii)}{=} \frac{n}{2} \cdot \frac{2}{n} = 1\). Contradiction.

Case 2. \(n\) is odd. \(1 \geq P[(A_1 \cap A_2) \sqcup (A_3 \cap A_4) \sqcup \cdots \sqcup (A_{n-2} \cap A_1) \sqcup A_n] \overset{(ii)}{>} p + \frac{n-1}{2} \cdot \frac{2}{n} \geq \frac{1}{n} + \frac{n-1}{n} = 1\). Again contradiction.

Problem 2—Stat 401. Let \(\xi, \eta\) be two independent standard normal random variables. Find the PDF of \(\xi/\eta\).

Solution to Problem 2. Let \(F(x)\) be the CDF of \(\xi/\eta\). Therefore

\[
F(x) = P\left(\frac{\xi}{\eta} \leq x\right) = P\left(\frac{\xi}{\eta} \leq x; \eta > 0\right) + P\left(\frac{\xi}{\eta} \leq x; \eta < 0\right) = P(\xi \leq x\eta; \eta > 0) + P(\xi \geq x\eta; \eta < 0) = 2P(\xi \leq x\eta; \eta > 0)
\]

\[
= 2 \int_{t \leq xs, s > 0} \frac{1}{2\pi} e^{-\frac{x^2+s^2}{2}} ds dt
\]

\[
= \frac{1}{\pi} \int_0^\infty ds \int_{-\infty}^{xs} e^{-\frac{s^2+x^2}{2}} dt.
\]

Taking derivatives in \(x\) yields

\[
f(x) := F'(x) = \frac{1}{\pi} \int_0^\infty se^{-\frac{1}{2}(1+x^2)} ds = \frac{1}{\pi(1+x^2)}, \text{ for all } x \in \mathbb{R}.
\]

Problem 3—Stat 411. Let \(X_{(1)} < X_{(2)} < \ldots < X_{(n)}\) be the order statistics of a random sample of size \(n\) from the exponential distribution with pdf

\[
f(x|\theta_1, \theta_2) = \frac{1}{\theta_2} e^{-(x-\theta_1)/\theta_2}, \quad x \geq \theta_1
\]
(i) Derive the MLE of \((\theta_1, \theta_2)\).

(ii) Show that \(Z_2 = (n - 1)(X_2 - X_1)\), \(Z_3 = (n - 2)(X_3 - X_2)\), \ldots, \(Z_n = X_n - X_{n-1}\) are independent and each \(Z_i\) has identical exponential distribution.

**Solution to Problem 3.**

(i) The joint distribution of \(X_1 < X_2 < \ldots < X_n\) is

\[
g(x(1), \ldots, x(n)) = \frac{n!}{\theta_2^n} \exp\left\{-\frac{\sum_{i=1}^n (x(i) - n\theta_1)}{\theta_2}\right\} I_{x(1) \geq \theta_1, x(i+1) > x(i), i=1, \ldots, n}.
\]

For given \(\theta_1\), \(g(x(1), \ldots, x(n))\) is maximized when \(\theta_2 = \sum_{i=1}^n (x(i) - n\theta_1) / n\) and the maximum value is

\[
\frac{n!}{(\sum_{i=1}^n (x(i) - n\theta_1)/n)^n} \exp(-n) I_{x(1) \geq \theta_1, x(i+1) > x(i), i=1, \ldots, n}.
\]

The maximum value is an increasing function of \(\theta_1\). Since \(\theta_1 \leq X_1\), then it will be maximized at \(\theta_1 = X_1\). The MLE of \((\theta_1, \theta_2)\) is \((\hat{\theta}_1 = X_1, \hat{\theta}_2 = \sum_{i=1}^n (X(i) - nX_1(n))/n)\).

(ii) The joint distribution of \(X_1 < X_2 < \ldots < X_n\) is given above. Let \(Z_1 = nX_1(1)\). Then we have

\[
X_1 = \frac{Z_1}{n}, \quad X_2 = \frac{Z_2}{n-1} + \frac{Z_1}{n}, \quad X_3 = \frac{Z_3}{n-2} + \frac{Z_2}{n-1} + \frac{Z_1}{n}, \quad \ldots, \quad X_n = \frac{Z_n}{n-1} + \frac{Z_{n-1}}{n-2} + \frac{Z_1}{n}.
\]

The Jacobian matrix is

\[
J = \begin{pmatrix}
\frac{\partial X_1}{\partial Z_1} & \frac{\partial X_1}{\partial Z_2} & \cdots & \frac{\partial X_1}{\partial Z_n} \\
\frac{\partial X_2}{\partial Z_1} & \frac{\partial X_2}{\partial Z_2} & \cdots & \frac{\partial X_2}{\partial Z_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial X_n}{\partial Z_1} & \frac{\partial X_n}{\partial Z_2} & \cdots & \frac{\partial X_n}{\partial Z_n}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{n} & 0 & \cdots & 0 \\
\frac{1}{n} & \frac{1}{n-1} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{n} & \frac{1}{n-1} & \cdots & 1
\end{pmatrix} = \frac{1}{n!}
\]

Therefore the joint distribution of \(Z_1, \ldots, Z_n\) is

\[
f(z_1, \ldots, z_n) = |J| g(x(1), \ldots, x(n)) = \frac{1}{\theta_2^n} \exp\left\{-\frac{\sum_{i=1}^n (z_i - n\theta_1)}{\theta_2}\right\} I_{z_1 \geq n\theta_1, z_i > 0, i=2, \ldots, n}.
\]

On the other hand, the density function of \(X_1(1)\) is

\[
n \frac{1}{\theta_1} e^{-\frac{n\theta_1 - n\theta_1}{\theta_2}} I_{X_1(1) \geq \theta_1}
\]

Thus the density function of \(Z_1\) is \(f(z_1) = \frac{1}{\theta_1} e^{-\frac{z_1-n\theta_1}{\theta_2}} I_{z_1 \geq n\theta_1}\). Consequently we have

\[
f(z_1, \ldots, z_n) = f(z_1) \frac{1}{\theta_2^{n-1}} \exp\left\{-\frac{\sum_{i=2}^n z_i}{\theta_2}\right\} I_{z_i > 0, i=2, \ldots, n}.
\]

This clearly shows that \(Z_2, \ldots, Z_n\) are independent with exponential distribution.
Problem 4—Stat 411. Let $X_1, X_2, \ldots, X_n$ denote a random sample from the normal distribution $N(\theta, 1)$, where $-\infty < \theta < \infty$.

(a) Show that $Y = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic for $\theta$.

(b) Determine the MVUE of $\theta^2$.

(c) If a constant $b$ is defined by the equation $P(X_1 \leq b) = 0.95$. Find the MVUE of $b$.

(For your reference, $P(Z \leq 1.645) = 0.95$ if $Z \sim N(0, 1)$.)

Solution to Problem 4.

(a) The density function of $X_i$ is

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x - \theta)^2\right\}$$

with $x \in (-\infty, \infty)$, which belongs to the regular exponential class with $K(x) \equiv x$. Therefore,

$$Y = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} K(X_i)$$

is complete and sufficient for $\theta$.

(b) Since $X_i \text{i.i.d.} \sim N(\theta, 1)$ implies $Y \sim N(n\theta, n)$, then

$$E(Y^2) = \text{Var}(Y) + [E(Y)]^2 = n + (n\theta)^2 = n + n^2\theta^2$$

Let $T = (Y^2 - n)/n^2$. Then $E(T) = \theta^2$.

Since $Y$ is complete sufficient and $T$ is a function of $Y$ which is unbiased for $\theta^2$, then $T$ is the MVUE of $\theta^2$.

(c) Since $0.95 = P(X_1 \leq b) = P(X_1 - \theta \leq b - \theta)$, then $b - \theta = 1.645$.

Let $U = Y/n + 1.645$ which is a function of $Y$. Since $E(U) = n\theta/n + 1.645 = \theta + 1.645 = b$, then $U$ is the MVUE of $b$.

Problem 5—Stat 411. Let $X_1, \ldots, X_n$ be iid samples from an exponential distribution with density function

$$f_\theta(x) = (1/\theta)e^{-x/\theta}, \quad x > 0, \quad \theta > 0.$$ 

The goal is to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0$ is some fixed value.
1. Suppose that the original data are corrupted somehow and all that you are left with are the binary data $Y_1, \ldots, Y_n$, where
\[
Y_i = I_{X_i \leq c} = \begin{cases} 
1 & \text{if } X_i \leq c \\
0 & \text{if } X_i > c,
\end{cases} \quad i = 1, \ldots, n
\]
and $c > 0$ is some fixed cutoff, e.g., $c = 7$. How would you test $H_0$ versus $H_1$ based on only these corrupted data? Give sufficient details about the null distribution that you would use to make the test achieve a given size.

2. It is intuitively clear that the corrupted data is not as good as the original data. Make this intuition precise by explaining how the power of your test based on the corrupted data compares to the power of a test based on the original data.

Solution to Problem 5.

1. It is easy to see that $Y_1, \ldots, Y_n$ are iid $\text{Ber}(p)$, where
\[
p = p_\theta = P_\theta(X_1 \leq c) = 1 - e^{-c/\theta}.
\]
Since $p_\theta$ is a decreasing function of $\theta$, the stated hypothesis testing problem is equivalent to testing
\[
H_0' : p = p_\theta_0 \quad \text{vs.} \quad H_1' : p < p_\theta_0
\]
based on the corrupted data. For this problem, the most powerful test is of the form
\[
\text{reject } H_0' \text{ iff } \sum_{i=1}^{n} Y_i < k,
\]
where $k$ is chosen to control the size of the test. A conservative test could be derived based on the binomial distribution of $\sum_{i=1}^{n} Y_i$ under the null; alternatively, if $n$ is reasonably large, an approximate test can be derived based on the normal approximation to binomial.

2. There is a uniformly most powerful test based on the original data. The test based on the corrupted data is, itself, a test based on the original data—given the original data, we could always corrupt it ourselves, if desired. So the Neyman–Pearson level implies that the most powerful test based on original data is more powerful than the test based on the corrupted data; therefore, the corrupted data is not as good.

Problem 6—Stat 416. A researcher is interested in learning if a new drug is better than a placebo in treating a certain disease. Because of the nature of the disease, only a limited number of patients can be found. Out of these, five are randomly assigned to the placebo and five to the new drug. Suppose that the concentration of a certain chemical in blood is measured and smaller measurements are better and the data are:
Drug: 3.2, 2.1, 2.3, 1.2, 1.5
Placebo: 3.4, 3.5, 4.1, 1.7, 2.2
(a) Let random variables $X$ and $Y$ stand for the output of a randomly selected person in the Drug group and Placebo group, respectively. Then the goal is to test if $X$ is \textit{stochastically smaller} than $Y$. Interpret the definition of “stochastically smaller”.

(b) Calculate the Mann-Whitney $U$ Test statistic on the given dataset. Specify the type of rejection region (that is, reject $H_0$ if $U$ is too small/large/small or large).

(c) Name another test for the same purpose (no need to perform it). Compared with the Mann-Whitney $U$ Test in (a), which test do you prefer? Explain your reason(s).

\textbf{Solution to Problem 6.}

(a) $X$ is \textit{stochastically smaller} than $Y$ if $F_X(x) \geq F_Y(x)$ for all $x$ and $F_X(x) > F_Y(x)$ for some $x$. Here $F_X$ and $F_Y$ stand for the cumulative distribution functions of $X$ and $Y$, respectively.

(b) The Mann-Whitney $U = \# \{(i,j) \mid X_i > Y_j\} = 5$. We reject $H_0$ if $U$ is too small.

(c) Other tests include Wald-Wolfowitz runs test, K-S two-sample test, median test, control median test, etc. Mann-Whitney $U$ test is far preferable as a test of location since it is more powerful.

\textbf{Problem 7—Stat 431.} For a sample of size $n = 2$, let $\pi_i$ be the desired probability of inclusion for sample $i$, with the usual constrain that $\sum_{i=1}^{N} \pi_i = n$. Let $\psi_i = \pi_i/2$ and $a_i = \frac{\psi_i (1-\psi_i)}{1-\pi_i}$. Draw the first sample with probability $a_i / \sum_{k=1}^{N} a_k$ of selecting sample $i$. Supposing sample $i$ is selected at the first draw, select the second sample $j$ from the remaining $N-1$ samples with probability $\psi_j / (1-\psi_i)$.

(i) Show that the probability that samples $i$ and $j$ are in sample is

$$\pi_{ij} = \frac{\psi_i \psi_j}{\sum_{k=1}^{N} a_k} \left( \frac{1}{1 - \pi_i} + \frac{1}{1 - \pi_j} \right)$$

(ii) Show that $P(\text{sample } i \text{ selected in sample}) = \pi_i$.

\textbf{Solution to Problem 7.}

(i)

$$P(\text{samples } i \text{ and } j \text{ are in sample})$$

$$= P(\text{samples } i \text{ drawn first and } j \text{ drawn second})$$

$$+ P(\text{samples } j \text{ drawn first and } i \text{ drawn second})$$

$$= \frac{a_i}{\sum_{k=1}^{N} a_k} \frac{\psi_j}{1 - \psi_i} + \frac{a_j}{\sum_{k=1}^{N} a_k} \frac{\psi_i}{1 - \psi_j}$$

$$= \frac{\psi_i (1-\psi_i) \psi_j}{\sum_{k=1}^{N} a_k (1 - \psi_i)(1 - \pi_i)} + \frac{\psi_j (1-\psi_j) \psi_i}{\sum_{k=1}^{N} a_k (1 - \psi_j)(1 - \pi_j)}$$

$$= \frac{\psi_i \psi_j}{\sum_{k=1}^{N} a_k} \left( \frac{1}{1 - \pi_i} + \frac{1}{1 - \pi_j} \right)$$
(ii)

\[ P(\text{sample } i \text{ in sample}) = \sum_{j=1, j \neq i}^{N} \pi_{ij} \]

\[ = \sum_{j=1, j \neq i}^{N} \frac{\psi_i \psi_j}{\sum_{k=1}^{N} a_k} \left( \frac{1}{1 - \pi_i} + \frac{1}{1 - \pi_j} \right) \]

\[ = \frac{\psi_i}{\sum_{k=1}^{N} a_k} \left( \frac{N}{\sum_{j=1}^{N} \psi_j} \left( \frac{1}{1 - \pi_i} + \frac{1}{1 - \pi_j} \right) - 2 \frac{\psi_i}{1 - \pi_i} \right) \]

\[ = \frac{\psi_i}{\sum_{k=1}^{N} a_k} \left( \frac{1}{1 - \pi_i} + \sum_{j=1}^{N} \frac{\psi_j - \pi_i}{1 - \pi_j} \right) \]

On the other hand, we can also show that

\[ 2 \sum_{k=1}^{N} a_k = \sum_{k=1}^{N} \psi_k + \psi_k (1 - 2 \psi_k) \]

\[ = \sum_{k=1}^{N} \frac{\psi_k (1 - \pi_k)}{1 - \pi_k} + \sum_{k=1}^{N} \frac{\psi_k}{1 - \pi_k} \]

\[ = 1 + \sum_{k=1}^{N} \frac{\psi_k}{1 - \pi_k}. \]  

The conclusion follows by combining the preceding two equations.

**Problem 8—Stat 451.** Set \( D = \{(x, y) : |x| + |y| \leq 1\} \), and suppose that \((X, Y)\) has uniform distribution on \(D\), i.e., the joint density function satisfies

\[ f(x, y) = \begin{cases} 
\frac{1}{2}, & \text{if } (x, y) \in D, \\
0, & \text{otherwise.} 
\end{cases} \]

Write down a Gibbs sampler algorithm to simulate from \(f(x, y)\). (Hint: It may be helpful to sketch a picture of the region \(D\).)

**Solution to Problem 8.** The joint density is uniform, so the two conditional distributions—\(X\) given \(Y\) and \(Y\) given \(X\)—must also be uniform, as their densities are proportional to the joint. But the ranges depend on the variable being conditioned on. Specifically, we have

\[ X \mid Y = y \sim \text{Unif}(-(1 - |y|), 1 - |y|) \]

\[ Y \mid X = x \sim \text{Unif}(-(1 - |x|), 1 - |x|) \].

7
Simulation from these two conditional distributions is straightforward, e.g., we can use `runif` in R. Alternating between simulations from these two generates a sequence \(\{(X_t, Y_t) : t \geq 1\}\), a Markov chain, and the stationary distribution of the chain is the uniform distribution on the disc.

**Problem 9—Stat 461.** Show that a finite-state aperiodic irreducible Markov chain is regular.

**Solution to Problem 9.** We need to show that for any \(i, j\) there is a large enough \(M\) such that \(p_{ij}^{(n)} > 0\) for all \(n \geq M\). Now we fix any pair \(i\) and \(j\). Since the Markov chain is irreducible, we know there exist an integer \(k\) and a large enough integer \(N\) such that

\[
p^{k+nd(j)} > 0,
\]

for all \(n \geq N\), where \(d(j)\) is the period of state \(j\). Now recall that the Markov chain is also aperiodic, hence \(d(j) = 1\). The proof is then completed by setting \(M = k + N\).

**Problem 10—Stat 461.** Customers arrive at a facility according to a Poisson process \(X(t)\) with rate \(\lambda\). Each customer pays $1 on arrival, and it is desired to evaluate the expected value of the total sum collected during the time \((0, t]\) discounted back to time 0. This quantity is given by

\[
M = \mathbb{E}\left[\frac{X(t)}{\sum_{k=1}^{n} e^{-rW_k}}\right],
\]

where \(r\) is the discount rate, \(W_1, W_2, \ldots\) are the arrival times, and \(X(t)\) is the total number of arrivals in \((0, t]\). Evaluate \(M\) explicitly in terms of \(\lambda, r\) and \(t\).

**Solution to Problem 10.** By the law of total probability, we have

\[
M = \sum_{n=1}^{\infty} \mathbb{E}\left[\sum_{k=1}^{n} e^{-rW_k} | X(t) = n \right] \mathbb{P}\{X(t) = n\}. \tag{4}
\]

Let \(U_1, \ldots, U_n\) denote independent random variables that are uniformly distributed in \((0, t]\). Because of the symmetry of the functional \(\sum_{k=1}^{n} \exp\{-rW_k\}\), we have

\[
\mathbb{E}\left[\sum_{k=1}^{n} e^{-rW_k} | X(t) = n\right] = \mathbb{E}\left[\sum_{k=1}^{n} e^{-rU_k}\right]
= n\mathbb{E}e^{-rt} \\
= \frac{n}{rt}(1 - e^{-rt}).
\]
Substitution into (4) gives

\[ M = \frac{1}{rt}(1 - e^{-rt}) \sum_{n=1}^{\infty} nP\{X(t) = n\} \]

\[ = \frac{1}{rt}(1 - e^{-rt})E[X(t)] \]

\[ = \frac{\lambda}{r}(1 - e^{-rt}). \]

**Problem 11—Stat 481.** Observations: for \( k \) brands of beer, researchers recorded the sodium content of \( n \) 12-ounce bottles. Denote the measurement responses by \( Y_{ij}, i = 1,\ldots,k; j = 1,\ldots,n \). Questions of interest: what is the “grand mean” sodium content? How much variability is there from brand to brand?

(1). Given the research objective and observations, what is the statistical model and basic assumptions that you think appropriate for the analysis? What is your hypotheses on the study for the variability from brand to brand?

(2). Derive \( E(MSTR) \).

(3). Derive the sampling distribution of the overall sample mean \( \bar{Y}_n \) and construct the confidence interval for the grand mean \( \mu \).

**Solution to Problem 11.**

(1). One-Way random effect model \( Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, i = 1,\ldots,k; j = 1,\ldots,n \), where \( \alpha_i \sim^{i.i.d.} N(0, \sigma^2_\alpha) \) and \( \varepsilon_{ij} \sim^{i.i.d.} N(0, \sigma^2) \) are two independent random components in the model. Hypotheses for variability is \( H_0 : \sigma^2_\alpha = 0 \) vs \( H_1 : \sigma^2_\alpha \neq 0 \).

(2). Note that \( \alpha_i \sim^{i.i.d.} N(0, \sigma^2_\alpha) \), \( \varepsilon_{ij} \sim^{i.i.d.} N(0, \sigma^2/n) \),

\[ E(SSTR) = E\left( \sum_{i=1}^{k} \sum_{j=1}^{n} (\bar{Y}_i - \bar{Y}_n)^2 \right) = E\left( \sum_{i=1}^{k} \sum_{i=1}^{n} (\alpha_i + \bar{\varepsilon}_i - \bar{\alpha} - \bar{\varepsilon})^2 \right) \]

\[ = n \left[ E \sum_{i=1}^{k} (\alpha_i - \bar{\alpha})^2 + E \sum_{i=1}^{k} (\bar{\varepsilon}_i - \bar{\varepsilon})^2 \right] \]

\[ = n \left[ (k-1)\sigma^2_\alpha + (k-1) \frac{\sigma^2}{n} \right] = n(k-1)\sigma^2_\alpha + (k-1)\sigma^2, \]

\[ E(MSTR) = E(SSTR/ (k-1)) = n\sigma^2_\alpha + \sigma^2. \]

(3). The overall sample mean overall

\[ \bar{Y}_n = \mu + \bar{\alpha} + \bar{\varepsilon}_n \sim N\left( \mu, \frac{\sigma^2_\alpha}{k} + \frac{\sigma^2}{kn} \right) = N\left( \mu, \frac{\sigma^2 + n\sigma^2_\alpha}{kn} \right). \]

From (2), \( E(MSTR) = n\sigma^2_\alpha + \sigma^2 \), and the sampling distribution of \( \bar{Y}_n \) is

\[ \frac{\bar{Y}_n - \mu}{\sqrt{MSTR/(kn)}} \sim t(k-1) \]
as $DF (SSTR) = k - 1$. Then the confidence interval for grand mean $\mu$

\[ \bar{Y} \pm t_{\alpha/2} (k - 1) \sqrt{MSTR/(kn)}. \]

**Problem 12—Stat 481.** In forestry, the diameter of a tree at breast height (which is fairly easy to measure) is used to predict the height of a tree (a difficult measurement to obtain). Silviculturists working in British Columbia’s boreal forest conducted a series of spacing trials to predict the heights of several species of trees. The data set whitespruce.txt contains the breast height diameters (in centimeters, $x_i$) and heights (in meters, $Y_i$) for a random sample of 36 white spruce trees.

(1). Consider a linear regression model, write down the model and assumptions.

(2). Based on the least square estimation, $\hat{\beta}_0 = 9.15, \hat{\beta}_1 = 0.48$ and standard errors are $se (\hat{\beta}_0) = 1.12, se (\hat{\beta}_1) = 0.06$. Is there sufficient evidence to conclude that there is a linear association between breast height diameter and tree height? [$t_{0.005 (34)} = 2.73, t_{0.025 (34)} = 2.02$].

(3). Derive $E (SSR)$ given that $\hat{\beta}_1 \sim N (\beta_1, \sigma^2/S_{xx})$, where $\sigma^2$ is the error variance, and $S_{xx} = \sum (x_i - \bar{x})^2$.  

(4). Complete ANOVA table given that $SSE = 95.7, Sxx = 795.3$. Will you reach the same decision as in (2)? [$F_{0.01 (1, 34)} = 7.44, F_{0.05 (1, 34)} = 4.13$].

**Solution to Problem 12.** (1). Consider simple linear regression $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where i.i.d. errors $\epsilon_i \sim N (0, \sigma^2), i = 1, ..., n$.

(2). To test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$, under $H_0$,

\[ t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{se (\hat{\beta}_1)} \sim t (n - 2), t_o = \frac{0.48}{0.06} = 8 > t_{0.005 (34)} = 2.73 \]

which implies that p-value $< 0.01$, hence reject $H_0$. Or construct confidence interval for $\beta_1, \hat{\beta}_1 \pm t_{\alpha/2} (n - 2) se (\hat{\beta}_1)$, so 95% C.I. for $\beta_1$ is $0.48 \pm 2.02 * 0.06 = (0.36, 0.60)$, which doesn’t include 0.

(3). The predicted value is $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{Y} + \hat{\beta}_1 (x_i - \bar{x})$, it leads to

\[ SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = \beta_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 = \beta_1^2 \cdot S_{xx}, \]

\[ E (SSR) = E (\hat{\beta}_1^2) S_{xx} = (\beta_1^2 + \frac{\sigma^2}{S_{xx}}) S_{xx} = \beta_1^2 S_{xx} + \sigma^2 \]

as $E (\hat{\beta}_1^2) = E^2 (\hat{\beta}_1) + Var (\hat{\beta}_1), \hat{\beta}_1 \sim N (\beta_1, \sigma^2/S_{xx})$.

(4). We have $SSR = \hat{\beta}_1^2 \cdot S_{xx} = 0.48^2 * 795.3 = 183.24$. ANOVA table
We can reach the same decision of rejecting $H_0$ as in (2).

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>$p$ – value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>183.24</td>
<td>1</td>
<td>183.24</td>
<td>65.1</td>
<td>$&lt; 0.01$</td>
</tr>
<tr>
<td>Error</td>
<td>95.7</td>
<td>34</td>
<td>2.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>278.94</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>