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.

(a) Suppose $X$ has gamma distribution with parameters $\alpha$ and $\beta$ (both $\alpha$ and $\beta$ are positive), i.e., $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}, x > 0$. Derive the moment generating function $M(t)$ of $X$.

(b) Utilizing the conclusion in (a), prove that (i) the distribution of $X/\beta$ is still gamma distributed, where $X$ has gamma distribution with parameters $\alpha$ and $\beta$ and (ii) $\sum_{i=1}^{n} X_i$ is also gamma distributed where $X_1, \ldots, X_n$ denote a random sample from a gamma distribution with parameters $\alpha$ and $\beta$.

Solution to Problem 1. (a)

$$M(t) = E(e^{tX}) = \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta} dx$$

Let $y = x(1-\beta t)/\beta$. When $t < \frac{1}{\beta}$, we have

$$M(t) = \int_0^\infty \frac{\beta/(1-\beta t)}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y} dy$$

$$= \frac{1}{(1-\beta t)^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy$$

$$= \frac{1}{(1-\beta t)^\alpha}.$$

(b) The moment generating function of $X/\beta$ can be written as

$$E(e^{tX/\beta}) = M(t/\beta) = \frac{1}{(1-\beta t/\beta)^\alpha} = \frac{1}{(1-t)^\alpha}.$$  

Thus $X/\beta$ has Gamma distribution with $\alpha$ and $\beta = 1$.

The moment generating function of $\sum_{i=1}^{n} X_i$ can be written as

$$E(e^{t\sum_{i=1}^{n} X_i}) = \prod_{i=1}^{n} E(e^{tX_i}) = (M(t))^n = \frac{1}{(1-\beta t)^{n\alpha}}.$$  

Thus $\sum_{i=1}^{n} X_i$ has Gamma distribution with parameters $n\alpha$ and $\beta$.

Problem 2—Stat 401. Let $X$ and $Y$ be independent chi-square random variables with $r_1$ and $r_2$ degrees of freedom, respectively. Namely, $X \sim \Gamma(\alpha = r_1/2, \beta = 2)$, and $Y \sim \Gamma(\alpha = r_2/2, \beta = 2)$. Let $Z := \frac{X/r_1}{Y/r_2}$. The distribution of $Z$ is called an $F$-distribution. Derive the probability density function $f(z)$ of $Z$ by answering the following questions. Specify the support of the PDF’s in all the three questions.
1. Write down the joint PDF \( f(x, y) \) of \( X \) and \( Y \).

2. Let \( W := Y \). Find the joint PDF \( f(z, w) \) of \( Z \) and \( W \).

3. Find the PDF of \( Z \).

**Solution to Problem 2.**

1. By the independence of \( X \) and \( Y \),
\[
f(x, y) = f(x)f(y) = \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)}y^{r_2/2-1}e^{-(x+y)/2}, x, y > 0.
\]

2. \[
\begin{align*}
Z &= \frac{r_2}{r_1}X \\
W &= Y
\end{align*}
\implies \begin{cases} 
X = \frac{r_1}{r_2} \cdot ZW \\
Y = W,
\end{cases}
\]
So the Jacobian matrix of the transformation is
\[
J = \begin{pmatrix}
\frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\
\frac{\partial y}{\partial z} & \frac{\partial y}{\partial w}
\end{pmatrix} = \begin{pmatrix}
\frac{r_1}{r_2} & 0 \\
0 & 1
\end{pmatrix}
\]
Therefore the joint PDF of \( Z \) and \( W \) is
\[
f(z, w) = \frac{\left(\frac{r_1}{r_2}\right)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}}z^{r_1/2-1}w^{(r_1+r_2)/2-1}e^{-\frac{w}{2}\left(\frac{r_2}{r_1}+1\right)}, w, z > 0.
\]

3. By integrating \( f(z, w) \) with respect to \( w \), we get
\[
f(z) = \int_0^\infty f(z, w)dw = \frac{\left(\frac{r_1}{r_2}\right)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}}z^{r_1/2-1}w^{(r_1+r_2)/2-1}e^{-\frac{w}{2}\left(\frac{r_2}{r_1}+1\right)}dw
\]
\[
= \frac{\left(\frac{r_1}{r_2}\right)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}}2^{(r_1+r_2)/2}\left(\frac{r_1}{r_1z + r_2}\right)^{(r_1+r_2)/2}\Gamma\left(\frac{r_1 + r_2}{2}\right)
\]
\[
= \frac{\left(\frac{r_1}{r_2}\right)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}}z^{r_1/2-1}\left(\frac{r_1}{r_1z + r_2}\right)^{(r_1+r_2)/2}\Gamma\left(\frac{r_1 + r_2}{2}\right), \text{ when } z > 0.
\]

**Problem 3—Stat 411.** Let \( X_1, \ldots, X_n \) be a random sample from \( X \sim \text{Exp}(\theta) \), i.e.
\[
f(x, \theta) = \frac{1}{\theta}e^{-x/\theta}, x > 0,
\]
where \( \theta > 0 \).

1. Find the maximum likelihood estimator \( \hat{\theta}_{\text{mle}} \) for \( \theta \).

2. Find the Fisher information for \( \theta \). Is the \( \hat{\theta}_{\text{mle}} \) an efficient estimator?

3. What is the maximum likelihood estimator \( \hat{\theta}_{\text{mle}} \) for \( \theta \) given that \( \theta > \frac{1}{2} \).

3
**Solution to Problem 3.** (1). Likelihood function \( L(\theta) = \theta^{-n}e^{-\sum x_i/\theta} \). Likelihood equation

\[
\frac{\partial \log L(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum x_i}{\theta^2} = 0 \Rightarrow \hat{\theta} = \bar{X}
\]

(2). Mean and variance

\[
EX = \int_{0}^{\infty} x f(x) dx = \theta, \quad EX^2 = \int_{0}^{\infty} x^2 f(x) dx = 2\theta^2.
\]

It is an unbiased estimator, \( E(\bar{X}) = \theta \).

Fisher information for \( \theta \) is

\[
I(\theta) = -E\left[\frac{\partial^2 \log f(X, \theta)}{\partial \theta^2}\right] = \frac{1}{\theta^2} - 2\frac{1}{\theta^3} = \frac{1}{\theta^2},
\]

variance of the estimator is \( \text{Var}(\bar{X}) = \frac{\theta^2}{n} = \left[\frac{nI(\theta)}{n}\right]^{-1} \). It reaches the Rao-Cramer Lower Bound, i.e. the mle is an efficient estimator. thus \( \hat{\theta} \).

(3). Given that \( \theta > \frac{1}{2} \), then the maximum likelihood estimator is

\[
\hat{\theta}_{\text{mle}} = \begin{cases} 
\bar{X}, & \bar{X} < \frac{1}{2} \\
\frac{1}{2}, & \bar{X} \geq \frac{1}{2} = \min \left\{ \bar{X}, \frac{1}{2} \right\}
\end{cases}
\]

**Problem 4—Stat 411.** Let \( X_1, \ldots, X_n \) denote a random sample from an exponential distribution, i.e., \( f(x) = \frac{1}{\theta}e^{-x/\theta}, x > 0 \). Let \( H_0 : \theta = 1 \) and \( H_1 : \theta > 1 \).

(a) Show that there exists a uniformly most powerful test for \( H_0 \) against \( H_1 \). Determine the statistic \( Y \) upon which the test may be based, and derive the best critical region of size \( \alpha \).

(b) Find the distribution of the statistic \( Y \) in part (a). If we want a significance level of 0.05, write an equation which can be used to determine the critical region. Let \( \gamma(\theta), \theta \geq 1 \), be the power function of the test. Derive the expression of \( \gamma(\theta) \).

**Solution to Problem 4.** (a) The likelihood function can be written as

\[
L(\theta) = \exp(-n \log \theta - \sum_{i=1}^{n} X_i/\theta).
\]

For \( \theta > 1 \), we have

\[
\frac{L(1)}{L(\theta)} = \exp(n \log \theta + (1/\theta - 1) \sum_{i=1}^{n} X_i).
\]

So \( L(\theta) \) has monotone decreasing likelihood ratio in the statistics \( Y = \sum_{i=1}^{n} X_i \). Thus the UMP level \( \alpha \) critical region for testing \( H_0 \) versus \( H_1 \) is given by

\[
Y = \sum_{i=1}^{n} X_i \geq c,
\]

4
where $c$ is chosen such that $\alpha = P_{\theta=1}[Y \geq c]$.

(b) Since $X_i$ is a gamma distribution with $\alpha = 1$ and $\beta = \theta$, and $X_1, \ldots, X_n$ are independent, $Y = \sum_{i=1}^{n} X_i$ follows gamma distribution with $\alpha = n$ and $\beta = \theta$. Under $H_0$, $Y$ follows gamma distribution with $\alpha = n$ and $\beta = 1$. Notice that $2Y$ follows gamma distribution with $\alpha = n$ and $\beta = 2$, which is also $\chi^2$ distribution with df=$2n$. Thus we can choose $c$ such that $P(\chi^2(2n) > 2c) = \alpha$.

The power function $\gamma(\theta)$ can be computed similarly

$$\gamma(\theta) = P_{\theta}[Y \geq c] = P_{\theta}[\frac{2Y}{\theta} \geq \frac{2c}{\theta}] = P_{\theta}[\chi^2(2n) \geq \frac{2c}{\theta}].$$

You can also write $\gamma(\theta)$ as an integral using the density function of $\gamma(n, \theta)$.

**Problem 5—Stat 411.** Let $X_1, \ldots, X_n$ be iid random variables, with mean $\mu$ and variance $\sigma^2$. Write $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ for the sample mean and sample variance, respectively.

1. Show that $S^2$ is an unbiased estimator of $\sigma^2$.
2. Show that $S^2$ is a consistent estimator of $\sigma^2$.
3. Consider a Poisson model, where the mean and variance both equal $\theta$. In this case, both $\bar{X}$ and $S^2$ are reasonable estimators of $\theta$, for example, both are unbiased and consistent. Is one estimator better than the other? Clearly present your case, justifying each claim.

**Solution to Problem 5.**

1. Rewrite the sample variance as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right).$$

Then $E(X_i^2) = \sigma^2 + \mu^2$ for each $i$, and $E(\bar{X}^2) = \sigma^2/n + \mu^2$. Putting this together, gives

$$E(S^2) = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} (\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2) \right\} = \cdots = \sigma^2,$$

so $S^2$ is an unbiased estimator of $\sigma^2$.

2. Start with the version of the sample variance from above, i.e.,

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right).$$
The law of large numbers says that $\overline{X} \to \mu$ in probability, and the continuous mapping theorem says $\overline{X}^2 \to \mu^2$ in probability. The law of large numbers also says that $\frac{1}{n} \sum_{i=1}^{n} X_i^2 \to \sigma^2 + \mu^2$ in probability. Writing $S^2$ one more time as

$$S^2 = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{n}{n-1} \overline{X}^2,$$

since $n/(n-1) \to 1$ as $n \to \infty$, it is clear that the right-hand side converges to $\sigma^2 + \mu^2 - \mu^2 = \sigma^2$ in probability. Therefore, $S^2 \to \sigma^2$ in probability, i.e., $S^2$ is a consistent estimator of $\sigma^2$.

3. Since $\overline{X}$ and $S^2$ are both unbiased and consistent, a natural way to compare them is to consider the variance; that is, if there is one with uniformly smaller variance, then it’s the better estimator. In this case, it is not easy to write a formula for the variance of the sample variance, so a direct comparison is difficult. However, we know that, for the Poisson model, the sample mean $\overline{X}$ is a complete sufficient statistic. The Lehmann–Scheffe theorem that if there exists an unbiased estimator that is a function of a complete sufficient statistic, then that must be the minimum variance unbiased estimator. Since $S^2$ is not a function of a complete sufficient statistic, while $\overline{X}$ is, it must be that the variance of $\overline{X}$ is uniformly smaller than that of $S^2$; therefore, the sample mean is a better estimator.

**Problem 6—Stat 416.** A time and motion study was made in the permanent mold department at Central Foundry to determine whether there was a pattern to the variation in the time required to pour the molten metal into the die and form a casting of a $6 \times 4$ in. Y-shaped branch. The metallurgical engineer suspected that pouring times before lunch were shorter than pouring times after lunch on a given day. Twelve independent observations in seconds were taken throughout the day, six before lunch and six after lunch.

Before Lunch: 12.6, 11.2, 11.4, 9.4, 13.2, 12.0
After Lunch: 16.4, 15.4, 14.1, 14.0, 13.4, 11.3

(a) Use Wilcoxon Rank-Sum test to find the $P$ value for the alternative that mean pouring time before lunch is less than after lunch for the data below on pouring times in seconds. Specify your null hypothesis and alternative hypothesis.

(b) It is known that “the asymptotic relative efficiency of the Wilcoxon Rank-Sum test relative to the two-sample Student’s $t$ test is 1.50 for the double exponential distribution and 1.09 for the logistic distribution, which are both heavy-tailed distributions”. Explain the concept “asymptotic relative efficiency” and the implication of this statement.

**Solution to Problem 6.** (a) With equal sample sizes $m = n = 6$, let $X$ denote pouring time before lunch and $Y$ denote pouring time after lunch. The null hypothesis is $H_0 : \theta = \mu_Y - \mu_X = 0$ and the desired alternative hypothesis is $H_1 : \theta > 0$ and the appropriate $P$ value is in the left tail for the Wilcoxon Rank-Sum test statistic $W_N$. 

6
The pooled array with $X$ values underlined is
$\begin{align*}
9.4, 11.2, 11.3, 11.4, 12.0, 12.6, 13.2, 13.4, 14.0, 14.1, 15.4, 16.4,
\end{align*}$
and $W_N = 1 + 2 + 4 + 5 + 6 + 7 = 25$. The $P$ value is $P(W_N \leq 25) = 0.013$ from Table J for $m = 6, n = 6$. Thus, the null hypothesis $H_0 : \theta = 0$ is rejected in favor of the alternative $H_1 : \theta > 0$ at any significance level $\alpha \geq 0.013$.

(b) Let $A$ and $B$ be two consistent tests of a null hypothesis $H_0$ and an alternative hypothesis $H_1$, at significance level $\alpha$. The asymptotic relative efficiency (ARE) of test $A$ relative to test $B$ is the limiting value of the ratio $n_b/n_a$, where $n_a$ is the number of observations required by test $A$ for the power of test $A$ to equal the power of test $B$ based on $n_b$ observations while simultaneously $n_b \to \infty$ and $H_1 \to H_0$.

The statement implies that the Wilcoxon Rank-Sum test is preferable to the two-sample $t$ test for heavy-tailed distributions since it is more powerful.

**Problem 7—Stat 431.** In a study to compare the precision of systematic and stratified sampling for estimating the average concentration of lead in the soil. The $1 \text{ km}^2$ area was divided into 100-m square, and a soil sample was collected at each of the resulting 121 grid intersections. Summary statistics from this systematic sample are given below:

<table>
<thead>
<tr>
<th>Element</th>
<th>$n$</th>
<th>Average (mg/kg)</th>
<th>Range (mg/kg)</th>
<th>Standard Deviation (mg/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead</td>
<td>121</td>
<td>127</td>
<td>22-942</td>
<td>146</td>
</tr>
</tbody>
</table>

The investigator also poststratified the same region. Stratum A considered of farmland away from roads, villages, and woodlands. Stratum B contained areas within 50m of roads, and was expected to have larger concentrations of lead. Stratum C contained the woodland, which also were expected to have larger concentration of lead because of foliage would capture airborne particles. The data on concentration of lead were not used in determining the strata. The data from the grid points falling in each stratum are in the following table:

<table>
<thead>
<tr>
<th>Element</th>
<th>Stratum</th>
<th>$n_b$</th>
<th>Average (mg/kg)</th>
<th>Range (mg/kg)</th>
<th>Standard Deviation (mg/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead</td>
<td>A</td>
<td>82</td>
<td>71</td>
<td>22-201</td>
<td>28</td>
</tr>
<tr>
<td>Lead</td>
<td>B</td>
<td>31</td>
<td>259</td>
<td>36-942</td>
<td>232</td>
</tr>
<tr>
<td>Lead</td>
<td>C</td>
<td>8</td>
<td>189</td>
<td>88-308</td>
<td>79</td>
</tr>
</tbody>
</table>

(a) Calculate a 95% for the average concentrate of lead in the area using systematic sample (Soil samples are collected at the grid intersections. We may assume the amount of soil is negligible compared with that in the region. You may also assume this sample behaves like an SRS).

(b) Now use the poststratified sample, and find 95% CIs for the average concentrate of lead. How do these compare with CI in (a)?

**Solution to Problem 7.** (a) By SRS formula, the sample mean is the estimation of population mean. An unbiased estimator of the variance of the estimation is given by

$$\sqrt{\left(1 - \frac{n}{N}\right)s^2} \frac{s^2}{n}.$$

Since the amount of soil is negligible compared with that in the region, we can ignore the finite population correction (fpc). Thus the 95% CI is

$$127 \pm 1.96 \frac{146}{\sqrt{121}} = [101.0, 153.0].$$
(b) Using the formula for computing poststratified estimator is given by
\[
\bar{y}_{\text{post}} = \sum_{h=1}^{H} \frac{N_h}{N} \bar{y}_h.
\]
Because samples are collected at the grid intersections, so we have \( \frac{N_h}{N} = \frac{n_h}{n} \). Thus
\[
\bar{y}_{\text{post}} = \frac{82}{121} \cdot 71 + \frac{31}{121} \cdot 259 + \frac{8}{121} \cdot 189 = 127.
\]
The estimation of the variance of \( \bar{y}_{\text{post}} \) is given by
\[
\hat{V}(\bar{y}_{\text{post}}) = \left(1 - \frac{n}{N}\right) \sum_{h=1}^{H} \frac{N_h s_h^2}{N n}.
\]
Based on the discussion above, we have
\[
\hat{V}(\bar{y}_{\text{post}}) = \frac{82}{121} \cdot \frac{28^2}{121} + \frac{31}{121} \cdot \frac{232^2}{121} + \frac{8}{121} \cdot \frac{79^2}{121} = 121.8.
\]
Thus the 95% CI is
\[
127 \pm 1.96 \frac{121.8}{\sqrt{121}} = [105.4, 148.6].
\]
Compared with SRS approach, poststratified approach increase the precision of the estimation.

**Problem 8—Stat 451.** Suppose that \((X,Y)\) has a bivariate normal distribution with zero means, unit variances, and correlation \(\rho \in (-1, 1)\). The joint density function is
\[
f_{X,Y}(x,y) \propto e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)}.
\]
1. Derive the conditional distribution of \(Y\), given \(X = x\).
2. Derive a Gibbs sampler for this bivariate normal.
3. Are the samples obtained by your Gibbs sampler exact bivariate normal samples, or are they approximate samples? If you think they are exact samples, then prove your claim; if you think they are approximate samples, then explain in what sense they are approximate.

**Solution to Problem 8.**
1. Take the conditional density of \(Y\), given \(X = x\), is proportional to the joint density as a function of \(y\) only:
\[
f_{Y|X}(y \mid x) \propto e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)}.
\]
It is easy to see that the conditional density must be a normal—exponential function with quadratic in the exponent. Moreover, for general normal distributions, the quadratic term in the density has coefficient \(-\frac{1}{2} \times \text{variance}\) and the linear term has coefficient mean/variance. So then it’s easy to check that the conditional distribution has mean \(\rho x\) and variance \(1 - \rho^2\).
2. Since the conditional distribution above is symmetric in the pair \((X,Y)\), the full conditionals are

\[ Y \mid (X = x) \sim N(\rho x, 1 - \rho^2) \quad \text{and} \quad X \mid (Y = y) \sim N(\rho y, 1 - \rho^2). \]

Therefore, a Gibbs sampler can be carried out by initializing \(X_0\) at, say, 0, and then iterating over \(t = 1, 2, \ldots\), as follows:

\[ Y_t \mid X_{t-1} \sim N(\rho X_{t-1}, 1 - \rho^2) \quad \text{and} \quad X_t \mid Y_t \sim N(\rho Y_t, 1 - \rho^2). \]

3. The output of the Gibbs sampler is not an exact sample from the bivariate normal. For one thing, the Gibbs output is a Markov chain, so there is some dependence. Furthermore, one cannot sample exactly from a joint distribution by looking only at the conditionals, that’s not how probability works; one can, however, sample from a marginal and a conditional. Anyway, the Gibbs output is not an exact sample. It is an approximation in the sense that, in a limit along the \(t\) sequence, the Markov chain converges to its stationary distribution, with is the bivariate normal, by construction. So, various expectations related to the bivariate normal can be accurately approximated based on a long sequence of Gibbs samples.

**Problem 9—Stat 461.** A Markov chain \(X_0, X_1, \ldots\) has the transition probability matrix

\[
P = \begin{pmatrix}
1 & 1 & 1 & 1 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & \frac{1}{2} & \frac{1}{2} \\
0 & 0 & \frac{1}{2} & \frac{1}{2}
\end{pmatrix}.
\]

Is state 0 transient or recurrent? Justify your answer.

**Solution to Problem 9.** A state \(i\) is recurrent if and only if

\[
\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty.
\]

So we only need to compute \(P_{00}^{(n)}\). But it is easy to see that

\[
P(n) = \begin{pmatrix}
\frac{1}{22^n} & \frac{1}{22^n} & * & * \\
\frac{1}{22^n} & \frac{1}{22^n} & * & * \\
0 & 0 & * & * \\
0 & 0 & * & *
\end{pmatrix}.
\]

Hence

\[
\sum_{n=1}^{\infty} P_{00}^{(n)} = \sum_{n=1}^{\infty} \frac{1}{22^n} < \infty.
\]

Therefore, state 0 is transient.
Problem 10—Stat 461.  This problem is related to probability generating function. It has two subquestions.

(a) Let $\xi$ and $\eta$ be independent nonnegative integer-valued random variables having the probability generating function

$$
\phi(s) = E s^\xi, \quad \text{and} \quad \psi(s) = E s^\eta, \quad |s| \leq 1.
$$

Show that the generating function for the random variable $\xi + \eta$ is $\phi(s)\psi(s)$.

(b) Let $\xi_1, \xi_2, ...$ be i.i.d. nonnegative integer-valued random variables with probability generating function $\phi(s) = E s^{\xi_1}$. Let $N$ be a nonnegative integer-valued random variable, independent of $\xi_1, \xi_2, ...$ with probability generating function $g(s) = E s^N$. Show that the probability generating function of the random sum

$$
X = \xi_1 + \cdots + \xi_N
$$

is

$$
E s^X = g(\phi(s)).
$$

Solution to Problem 10.

(a) By definition and independence of $\xi$ and $\eta$, we have

$$
E s^{(\xi + \eta)} = E e^{\xi} e^{\eta} = E s^{\xi} E s^{\eta} = \phi(s)\psi(s).
$$

(b) We have

$$
E s^X = \sum_{k=0}^{\infty} \Pr\{X + k\} s^k
$$

$$
= \sum_{k=0}^{\infty} \left( \sum_{n=0}^{\infty} \Pr\{X = k | N = n\} \Pr\{N = n\} \right) s^k
$$

$$
= \sum_{k=0}^{\infty} \left( \sum_{n=0}^{\infty} \Pr\{\xi_1 + \cdots + \xi_n = k | N = n\} \Pr\{N = n\} \right) s^k
$$

$$
= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \Pr\{\xi_1 + \cdots + \xi_n = k\} \Pr\{N = n\} s^k
$$

$$
= \sum_{n=0}^{\infty} \left( \sum_{k=0}^{\infty} \Pr\{\xi_1 + \cdots + \xi_n = k\} s^k \right) \Pr\{N = n\}
$$

$$
= \sum_{n=0}^{\infty} \phi(s)^n \Pr\{N = n\} \quad \text{[using (a)]}
$$

$$
= g(\phi(s)).
$$
Problem 11—Stat 481. Suppose you are studying the monthly amounts that seventh- and eighth-grade boys and girls spend on entertainment such as movies, music CDs, and candy. Representative samples of children within a certain school district were selected, and children were asked about their spending habits. The following results were obtained:

<table>
<thead>
<tr>
<th></th>
<th>7th-grade boys</th>
<th>8th-grade boys</th>
<th>7th-grade girls</th>
<th>8th-grade girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>30</td>
<td>25</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>Mean($)</td>
<td>20.1</td>
<td>23.2</td>
<td>19.6</td>
<td>25.0</td>
</tr>
<tr>
<td>Standard Deviation($)</td>
<td>6.0</td>
<td>5.6</td>
<td>5.3</td>
<td>7.0</td>
</tr>
</tbody>
</table>

(a) Test whether the four groups differ with respect to their mean spending amounts. Use significance level $\alpha = 0.05$ and $F$ critical value $F(0.05; 3, 106) = 2.69$.

(b) Test whether there is significant difference in the mean spending amounts of eighth-grade boys and eighth-grade girls. Use significance level $\alpha = 0.05$ and $t$ critical value $t(0.025; 24) = 2.06$.

Solution to Problem 11. (a)

$$SS_{Error} = (29)(6.0)^2 + (24)(5.6)^2 + (29)(5.3)^2 + (24)(7.0)^2 = 3787.25$$


$$F_{Treatment} = \frac{536.86/3}{3787.25/106} = 5.01 > 2.69$$

with $p$-value $< 0.05$. That is, there are significant differences among expenditures.

(b) We can use the 2-sample t-test with unequal variances.

$$T = \frac{25.0 - 23.2}{\sqrt{\frac{7.0^2}{25} + \frac{5.6^2}{25}}} = 1.00 < 2.06$$

That is, there is no significant difference in the mean spending amounts of eighth-grade boys and eighth-grade girls at 5% level.

Problem 12—Stat 481. A company operates children portrait studios in 21 cities. The manager is considering an expansion and wishes to investigate whether the sales ($Y$) in a community can be predicted from the number of persons aged 16 or younger ($x_1$) and per capita disposal personal income ($x_2$). Try to fit the data with a multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, i = 1, ..., n,$$

where the i.i.d. errors $\varepsilon_i \sim N(0, \sigma^2)$. The model can also be expressed in a matrix form $Y = X\beta + \varepsilon$.

(1) Complete the ANOVA table below for the linear model.
(2). State both null and alternative hypotheses you will test for the linear model. Draw your conclusion based on the ANOVA table.

<table>
<thead>
<tr>
<th>Source</th>
<th>S.S.</th>
<th>DF</th>
<th>M.S.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>24015</td>
<td>2</td>
<td>12007.5</td>
<td>99</td>
</tr>
<tr>
<td>Error</td>
<td>2180</td>
<td>18</td>
<td>121.1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26195</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Given: F(0.95,2,18)=3.55, F(0.95,2,19)=3.52, F(0.95,3,18)=3.15]

(3). Find the coefficient of determination $R^2$ and interpret it.

(4). Show that for the least square estimator $\hat{\beta}$, $\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$.

Solution to Problem 12. (1). ANOVA Table

(2). $H_0 : \beta_1 = \beta_2 = 0$ vs $H_1 : \text{at least one is not zero}$. $F = 99 > F(0.95, 2, 18) = 3.55$. Sales are related to the population and per capita disposal income.

(3). $R^2 = 91.7\%$, i.e. 91.7% of the variation in the data can be explained by the regression model above.

(4). Least square criterion: $\min_\beta \{Q(\beta)\} = \min_\beta \{(Y - X\beta)'(Y - X\beta)\}$

The the estimator $\hat{\beta}$ is the solution $\frac{\partial Q(\beta)}{\partial \beta} = 0$, i.e.

$$X'X\beta = X'Y \Rightarrow \hat{\beta} = (X'X)^{-1} X'Y$$

Thus

$$\text{Var}(\hat{\beta}) = \text{Var}\left((X'X)^{-1} X'Y\right)$$

$$= (X'X)^{-1} X' \cdot \text{Var}(Y) \cdot X (X'X)^{-1}$$

$$= (X'X)^{-1} X' \cdot \sigma^2 I_n \cdot X (X'X)^{-1} = \sigma^2 (X'X)^{-1}.$$