GROUND RULES:

- This exam contains 6 questions; each question is worth 10 points. The maximum number of points on this exam is 60.
- You may use a calculator if you wish, but SHOW ALL OF YOUR WORK AND EXPLAIN ALL OF YOUR REASONING!!!
- Tabled values for the standard normal, t, χ^2, and F distributions will be provided.
- *Any discussion or otherwise inappropriate communication between examinees, as well as the appearance of any unnecessary material, will be dealt with severely.*
- Print your name **at the top of this page in the upper right hand corner**.
- You have 3 hours to complete this exam. GOOD LUCK!

HONOR PLEDGE FOR THIS EXAM:

After you have finished the exam, please read the following statement and sign your name below it.

> *I promise that I did not discuss any aspect of this exam with anyone other than the instructor, that I neither gave nor received any unauthorized assistance on this exam, and that the work presented herein is entirely my own.*

1. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from the pdf

$$f_Y(y;\theta) = \begin{cases} \frac{2y}{\theta} e^{-y^2/\theta}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find a sufficient statistic for θ.
(b) Let U denote your sufficient statistic from part (a), and suppose that $\widehat{\theta}$ is any unbiased estimator of θ. Let

$$\widehat{\theta}^* = E(\widehat{\theta}|U).$$

Show that $\widehat{\theta}^*$ is unbiased and that $V(\widehat{\theta}^*) \leq V(\widehat{\theta})$. Essentially, I am asking you to prove the Rao-Blackwell Theorem here. *Hint:* Recall the iterated formulas for mean and variance from STAT 511. We also did this proof in class.
(c) What is the main implication of the result in part (b)? Talk about MVUE's and sufficiency. Note that you can answer this part even if you could not do part (b).

2. In an early-phase clinical trial, two drugs are administered to patients with hypertension. On each patient, we measure a point reduction of hypertension, which from past studies is known to follow a **normal distribution**. We have two independent samples

$$Y_{11}, Y_{12}, ..., Y_{1n_1} \text{ iid } \mathcal{N}(\mu_1, \sigma^2)$$
$$Y_{21}, Y_{22}, ..., Y_{2n_2} \text{ iid } \mathcal{N}(\mu_2, \sigma^2).$$

Physicians assume a **common variance** between the two drug populations. Here are some summary statistics from the trial (sample sizes, sample means, and sample standard deviations):

Drug 1	Drug 2
$n_1 = 11$	$n_2 = 13$
$\overline{y}_{1+} = 11.3$	$\overline{y}_{2+} = 14.6$
$s_1 = 6.3$	$s_2 = 7.7$

(a) Using this information, compute a 95 percent confidence interval for $\mu_1 - \mu_2$, the difference in the mean point reduction for the two drugs (under the common variance assumption). What does this confidence interval tell you about the two drugs?
(b) Suppose that a physician associated with the trial wanted to determine whether the two population variances were truly equal. Suggest a strategy on how she could accomplish this, and then execute your strategy with the information above.

3. The **Pareto distribution** is often used in economics to model income distributions. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a Pareto pdf of the form

$$f_Y(y;\theta) = \begin{cases} \theta\nu^\theta y^{-(\theta+1)}, & y > \nu \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 1$. In this model, ν represents the minimum income level (in \$1,000's). In this problem, we will assume that $\nu = 65$; i.e., ν is a known value. The parameter θ is unknown.

(a) Find a formula for the method of moments estimator for θ.
(b) Find a formula for the the maximum likelihood estimator for θ.
(c) Compute your MOM and MLE estimates of θ with the following data:

```
87.10  68.55  123.48  77.77  110.88  100.54  98.78  84.32
```

Your answers should be numerical.

4. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid Bernoulli(θ) sample; recall that the Bernoulli(θ) pmf is given by

$$f_Y(y;\theta) = \begin{cases} \theta^y(1-\theta)^{1-y}, & y = 0,1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the MVUE of θ.
(b) Find the MVUE of $\tau(\theta) = \theta^2$.

5. Prove **any two** of the following results.

- If $Y \sim \text{beta}(\theta,\theta)$, then $U = 1 - Y \sim \text{beta}(\theta,\theta)$.
- If $Y_1, Y_2, ..., Y_n$ are iid exponential with mean $\beta > 0$, then $Y_{(1)} \sim \text{exponential}(\beta/n)$.
- If $Y_1, Y_2, ..., Y_n$ are independent random variables, where $Y_i \sim \text{gamma}(\alpha_i, \beta)$, then $U = \sum_i Y_i \sim \text{gamma}(\sum_i \alpha_i, \beta)$.
- If $Y_1 \sim \text{gamma}(\alpha, 1)$, $Y_2 \sim \text{gamma}(\beta, 1)$, and Y_1 and Y_2 independent, then $U = Y_1/(Y_1 + Y_2) \sim \text{beta}(\alpha, \beta)$.

State explicitly which two you would like for me to grade. Proofs here should be mathematical.

6. In an agricultural experiment, researchers are interested in comparing three different fertilizers. The researchers decide to model their yields as follows:

$$\begin{aligned} &\text{Fertilizer 1:} \quad Y_{11}, Y_{12}, ..., Y_{1n_1} \sim \text{iid } \mathcal{N}(\mu_1, \sigma^2) \\ &\text{Fertilizer 2:} \quad Y_{21}, Y_{22}, ..., Y_{2n_2} \sim \text{iid } \mathcal{N}(\mu_2, \sigma^2) \\ &\text{Fertilizer 3:} \quad Y_{31}, Y_{32}, ..., Y_{3n_3} \sim \text{iid } \mathcal{N}(\mu_3, \sigma^2). \end{aligned}$$

The population means μ_1, μ_2, and μ_3 are unknown parameters. The common population variance σ^2 is also unknown. Fertilizers were randomly assigned to plots, so they assume that the samples are **independent**. The researchers are interested in the linear combination

$$\theta = a_1\mu_1 + a_2\mu_2 + a_3\mu_3,$$

where a_1, a_2, and a_3 are known constants. As an estimator for θ, they use

$$\widehat{\theta} = a_1\overline{Y}_{1+} + a_2\overline{Y}_{2+} + a_3\overline{Y}_{3+},$$

where $\overline{Y}_{i+}$ denotes the ith sample mean; $i = 1, 2, 3$.

(a) Argue that

$$\widehat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2\left[\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}\right]\right).$$

(b) Define

$$S_{p:3}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2}{n_1+n_2+n_3-3}.$$

Show that $(n_1+n_2+n_3-3)S_{p:3}^2/\sigma^2$ has a χ^2 distribution with $n_1+n_2+n_3-3$ degrees of freedom.

(c) Show that

$$t = \frac{\widehat{\theta} - \theta}{S_{p:3}\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}}$$

has a t distribution with $n_1+n_2+n_3-3$ degrees of freedom.

(d) Explain how to use the result in (c) to derive a $100(1-\alpha)$ percent confidence interval for θ.