Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Option B
August 30, 2016
12:30-4:30pm, Room 133 SMI

- There are a total of FOUR (4) problems in this exam. Please do all FOUR (4) problems.
- Each problem must be done in a separate exam book.
- Please turn in FOUR (4) exam books.
- Please write your code name and **NOT** your real name on each exam book.

1. Let $X_1, \ldots, X_n$ be an independent and identically distributed sample from the Beta(θ,1) distribution with population density

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

Please answer the following questions:

(a) Find the method of moments estimator of θ based on the first moment $\mu = \mathbb{E}_\theta(X_1)$, and find the maximum likelihood estimator of θ. Calculate their asymptotic relative efficiency.

(b) Suppose θ is equipped with a Gamma(a, b) prior. Calculate the Bayes estimator of θ under squared error loss.

(c) Demonstrate that a UMP level-α test exists for the hypotheses $H_0 : \theta \leq \theta_0$ vs. $H_A : \theta > \theta_0$. Calculate the power function of the test (the probability the test rejects H_0 as a function of θ) in terms of the cdf of a known distribution.

(d) Once again, suppose θ is equipped with a Gamma(a, b) prior. With H_0 and H_A as in part (c), derive the test that rejects when the posterior probability of H_0 is less than α.

(e) Exhibit a pivot based on the statistic $T = -\sum_{i=1}^{n} \log(X_i)$ and derive the shortest length $1 - \alpha$ confidence interval for θ based on that pivot.

2. A researcher makes observations Y_t, $t = 1, \cdots, 10$ and considers the following model:

$$Y_t = \alpha_t + \epsilon_t,$$

where $\epsilon_t, t = 1, \cdots, 10$ are independent and identically distributed with mean 0 and variance σ^2. Furthermore, $\alpha_t, t \geq 0$ is a piece-wise linear function given by

$$\alpha_t = \begin{cases} \beta_0 + \beta_1 t, & 0 \leq t \leq t^* \\ \beta_0^* + \beta_2 t, & t^* \leq t \leq 10. \end{cases}$$

The parameters β_0, β_1, β_0^*, and β_2 are unknown. However, since the function is known to be continuous, they satisfy a restriction that $\beta_0 + \beta_1 t^* = \beta_0^* + \beta_2 t^*$, for known $t^* = 5.5$.

(a) Write down a general linear model of the form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \epsilon$, where $\mathbf{b} = (b_0, b_1, b_2)^T$, and specify explicitly the vector $\mathbf{Y}$, the matrix $\mathbf{X}$, the covariance matrix of the error vector, and $\mathbf{b}$ in terms of β_0, β_1, β_0^*, and β_2.

(b) Obtain the normal equations for your model in (a).

(c) Find a linear unbiased estimator for $\beta_1 - \beta_2$ by solving the normal equations for $\beta_1 - \beta_2$ without necessarily inverting $X^T X$.

(d) Find the variance of your estimator for $\beta_1 - \beta_2$ from part (d).

(e) If the change point t^* was not known, would your model be still linear? Explain.

(f) Consider the following restriction for your model from (a): $\beta_2 - \beta_1 = 1$. Specify the normal equations for estimating β_0 and β_1.

3. *Parts (a), and (b) are independent and so you can work on part (b) separately from part (a).*

(a) Consider the model for a randomized complete block design: $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ where $i = 1, \ldots, K$, $j = 1, \ldots B$, and where α_i represents the effect of treatment i, β_j represents the effect of block j, and the ϵ_{ij} are all independent, with $N(0, \sigma_\epsilon^2)$. This is the unconstrained model for this design — no other assumptions about the α_i or β_j are imposed.

i. Consider the vector $\theta^T = (\mu, \alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_B)$. Let λ be a vector of length $K + B + 1$, assume $K = 2$ and $B = 3$. Explain what it means to say that $\lambda^T\theta$ is "estimable."

ii. For the model in part (a), is $\mu + \alpha_1 + \beta_1$ estimable? Explain.

iii. For the model in part (a), is $\mu + \alpha_1$ estimable? Explain.

iv. For the model in part (a), is $\alpha_1 - \alpha_2$ estimable? Explain.

(b) An experiment was conducted to study soil organic matter (SOM) concentrations in response to various cropping approaches. Specifically of interest was: (A) the use of organic fertilizer, chemical fertilizer, or a blend of organic and chemical fertilizer; (B) whether the cover crop was genetically modified (also call GM corn), conventional corn (non-GM corn), alfalfa, or wheat.

The experiment was conducted as follows. A large field was divided into 2 areas. Within each area, 12 plots were located, and at random, one of the combinations of fertilizer treatment and cover crop was assigned to each plot. This was done in such a way that each of the 12 combinations of fertilizer and cover crop appeared exactly once within each of the 2 areas.

The fertilizer was applied at the beginning of the growing season (in the spring) and each plot was planted according to the cover crop assigned to it. At the end of the growing season (in the fall) the investigators went to the center of each plot and removed some soil. They did this by digging a small hole to a depth of 30 cm, and removed about 100 gm of soil. Then, right next to that hole, they dug another hole to a depth of 70 cm, and removed 100 gm of soil from that hole. Finally, next to those two holes another was dug to 100 cm and 100 gm of soil were removed. Each of these samples of soil were analyzed in a standard way to measure soil organic matter.

The following output lists the data, some summary statistics including Type 1 Sums of Squares, and sample means for each group.

Please answer the following questions:

i. Construct the ANOVA table for this experiment, indicating source and degrees of freedom (df) only.

ii. Conduct a formal test for the presence of an interaction between cover crop and depth.

iii. Conduct a formal test for the presence of an interaction between cover crop and fertilizer.

iv. Consider the combination of chemical fertilizer and conventional corn. Estimate the mean soil organic matter for that combination, and provide a 90% confidence interval for that combination.

area	fert	crop	depth	som
1	org	gmcorn	30	66
1	org	gmcorn	70	57
1	org	gmcorn	100	57
1	org	convcorn	30	86
1	org	convcorn	70	80
1	org	convcorn	100	82
1	org	alf	30	87
1	org	alf	70	76
1	org	alf	100	77
1	org	wheat	30	96
1	org	wheat	70	85
1	org	wheat	100	87
1	chem	gmcorn	30	49
1	chem	gmcorn	70	42
1	chem	gmcorn	100	44
1	chem	convcorn	30	45
1	chem	convcorn	70	36
1	chem	convcorn	100	35
1	chem	alf	30	52
1	chem	alf	70	44
1	chem	alf	100	41
1	chem	wheat	30	50
1	chem	wheat	70	44
1	chem	wheat	100	45
1	blend	gmcorn	30	51
1	blend	gmcorn	70	45
1	blend	gmcorn	100	44
1	blend	convcorn	30	45
1	blend	convcorn	70	39
1	blend	convcorn	100	36
1	blend	alf	30	58
1	blend	alf	70	52
1	blend	alf	100	47
1	blend	wheat	30	48
1	blend	wheat	70	43
1	blend	wheat	100	42
2	org	gmcorn	30	75
2	org	gmcorn	70	62
2	org	gmcorn	100	63
2	org	convcorn	30	82
2	org	convcorn	70	72
2	org	convcorn	100	74
2	org	alf	30	95
2	org	alf	70	94
2	org	alf	100	91
2	org	wheat	30	97
2	org	wheat	70	86
2	org	wheat	100	83
2	chem	gmcorn	30	54
2	chem	gmcorn	70	43
2	chem	gmcorn	100	42
2	chem	convcorn	30	58
2	chem	convcorn	70	49
2	chem	convcorn	100	50

2	chem	alf	30	57
2	chem	alf	70	48
2	chem	alf	100	48
2	chem	wheat	30	56
2	chem	wheat	70	50
2	chem	wheat	100	45
2	blend	gmcorn	30	59
2	blend	gmcorn	70	50
2	blend	gmcorn	100	52
2	blend	convcorn	30	56
2	blend	convcorn	70	50
2	blend	convcorn	100	49
2	blend	alf	30	52
2	blend	alf	70	48
2	blend	alf	100	48
2	blend	wheat	30	44
2	blend	wheat	70	38
2	blend	wheat	100	36

The GLM Procedure

Dependent Variable: som

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	71	21968.31944	309.41295	.	.
Error	0	0.00000	.		
Corrected Total	71	21968.31944			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
area	1	284.01389	284.01389	.	.
fert	2	16922.19444	8461.09722	.	.
area*fert	2	40.86111	20.43056	.	.
crop	3	795.04167	265.01389	.	.
area*crop	3	126.81944	42.27315	.	.
fert*crop	6	1984.25000	330.70833	.	.
area*fert*crop	6	627.13889	104.52315	.	.
depth	2	1034.02778	517.01389	.	.
area*depth	2	1.36111	0.68056	.	.
fert*depth	4	19.97222	4.99306	.	.
area*fert*depth	4	13.63889	3.40972	.	.
crop*depth	6	15.41667	2.56944	.	.
area*crop*depth	6	50.30556	8.38426	.	.
fert*crop*depth	12	28.91667	2.40972	.	.
area*fert*crop*depth	12	24.36111	2.03009	.	.

The GLM Procedure
Least Squares Means

fert	crop	depth	som LSMEAN
blend	alf	30	55.0000000
blend	alf	70	50.0000000
blend	alf	100	47.5000000
blend	convcorn	30	50.5000000
blend	convcorn	70	44.5000000
blend	convcorn	100	42.5000000
blend	gmcorn	30	55.0000000
blend	gmcorn	70	47.5000000
blend	gmcorn	100	48.0000000
blend	wheat	30	46.0000000
blend	wheat	70	40.5000000
blend	wheat	100	39.0000000
chem	alf	30	54.5000000
chem	alf	70	46.0000000
chem	alf	100	44.5000000
chem	convcorn	30	51.5000000
chem	convcorn	70	42.5000000
chem	convcorn	100	42.5000000

chem	gmcorn	30	51.5000000
chem	gmcorn	70	42.5000000
chem	gmcorn	100	43.0000000
chem	wheat	30	53.0000000
chem	wheat	70	47.0000000
chem	wheat	100	45.0000000
org	alf	30	91.0000000
org	alf	70	85.0000000
org	alf	100	84.0000000
org	convcorn	30	84.0000000
org	convcorn	70	76.0000000
org	convcorn	100	78.0000000
org	gmcorn	30	70.5000000
org	gmcorn	70	59.5000000
org	gmcorn	100	60.0000000
org	wheat	30	96.5000000
org	wheat	70	85.5000000
org	wheat	100	85.0000000

4. For independent and identically distributed random variables $X_1, \ldots, X_n$, let $\mu = E(X_1)$ and $\sigma^2 = \text{var}(X_1) \in (0, \infty)$. Consider the random quantity $\mathcal{T}_n = \sqrt{n}\,\overline{X}/\widehat{\sigma}$, where $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}}$.

(a) Discuss the limit distribution (or behavior) of $\mathcal{T}_n$ as $n \to \infty$. Does your answer depend on μ?

(b) Define $\widetilde{\boldsymbol{u}}_n = (1, \ldots, 1)^T/\sqrt{n}$, $\boldsymbol{X} = (X_1, \ldots, X_n)^T$, and $\widetilde{\boldsymbol{X}} = \frac{\boldsymbol{X}}{\sqrt{\sum_{i=1}^n X_i^2}}$ to be $n \times 1$ vectors. Define by $\Theta \in [0, \pi]$ the angle between $\widetilde{\boldsymbol{u}}_n$ and $\boldsymbol{X}$, such that $\cos(\Theta) = \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{u}}_n$.

i. Derive the expression of $\mathcal{T}_n$ in terms of Θ.

ii. Discuss whether $\mathcal{T}_n$ is monotone increasing or decreasing with respect to Θ.

iii. Discuss whether $\mathcal{T}_n$ is monotone increasing or decreasing with respect to $\cos(\Theta)$.

iv. Consider the hypotheses $H_0 : \mu = 0$ versus $H_1 : \mu > 0$. Compute the p-value expressed in terms of the observed Θ.