MATH 668 EXAM 2 (April 25, 2016: 8:00am–10:30am) NAME ________________________

The exam is closed book; students are permitted to prepare one 8.5×11 page of formulas, notes, etc. that can be used during the exam. A calculator is permitted but not necessary for the exam.

Do 4 out of the 5 problems (20 points each, 80 points total). Clearly indicate the problem that you are omitting; if it is not clear, then the first 4 problems will be graded.

Problem 1. (20 points) Let A_t and B_t be the demand at time t for product A and the demand at time t for product B, respectively, and let G_t be the price of gasoline at time t. Andrea fits a simple linear regression model to predict the demand for product A based on the price of gasoline at that time; her fitted model based on the method of least squares is

$$(\text{Fitted value of } A_t) = 100 - 6G_t.$$

Boris fits a simple linear regression model to predict the demand for product B based on the price of gasoline at that time; his fitted model based on the method of least squares is

$$(\text{Fitted value of } B_t) = 60 - 5G_t.$$

Both Andrea and Boris also report their residuals so it is possible to regress either one's residuals on the other's residuals; this leads to two fitted models based on the method of least squares:

$$(\text{Fitted value of residual from Andrea's model}) = -0.2 \times (\text{Residual from Boris's model})$$

and

$$(\text{Fitted value of residual from Boris's model}) = -4 \times (\text{Residual from Andrea's model}).$$

Suppose Chris fits a linear regression model for predicting A_t based on G_t and B_t based on the method of least squares; i.e., Chris finds the values β_0, β_1, and β_2 which minimize $\sum_t (A_t - \beta_0 - \beta_1 G_t - \beta_2 B_t)^2$. Find the equation of the regression line in Chris's fitted model.

Problem 2. (20 points) Suppose that $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$ for $i = 1, \ldots, 4$ where the $x_1 = -2$, $x_2 = -1$, $x_3 = 1$, and $x_4 = 1$ are known values, the regression parameters β_0, β_1, and β_2 are unknown, and $e_1, \ldots, e_4$ are independent and identically distributed normal random variables with mean 0 and unknown variance σ^2.

Now, suppose that the x^2 term is excluded from the model, and least squares estimation is used to model the y's based on an intercept term and the x's (that is, suppose that we incorrectly use a linear model and find values, say γ_0 and γ_1, which minimize $\sum_{i=1}^{4} (y_i - \gamma_0 + \gamma_1 x_i)^2$). What is the bias of the least squares estimator of β_0? Write your answer as a function of the true (but unknown) parameter values β_0, β_1, and β_2.

Problem 3. (20 points) Consider the fixed effects model

$$y_{ij} \sim \text{independent Normal}(\mu_i, \sigma^2)$$

for $i = 1, \ldots, 3, j = 1, \ldots, 6$ where μ_1, μ_2, and μ_3 are fixed unknown constants.
(a - 10 pts) Using the fact that a Normal(μ, σ^2) density has the form

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\},$$

find the maximum likelihood estimators (MLEs) of μ_1, μ_2, μ_3, and σ^2 (as functions of the y_{ij}'s).
(b - 10 pts) Denote the MLEs of μ_i as $\hat{\mu}_i$ for $i = 1, 2, 3$. Let $R^2 = 1 - \frac{\sum_{i=1}^{3}\sum_{j=1}^{6}(y_{ij} - \hat{\mu}_i)^2}{\sum_{i=1}^{3}\sum_{j=1}^{6}(y_{ij} - \bar{y}_{\cdot\cdot})^2}$ where $\bar{y}_{\cdot\cdot} = \frac{1}{18}\sum_{i=1}^{3}\sum_{j=1}^{6} y_{ij}$. For what values of R^2 should the overall F-test of $H_0 : \mu_1 = \mu_2 = \mu_3$ (versus the alternative that H_0 is not true) be rejected at level .05?

Problem 4. (20 points) Suppose that

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \text{ for } i, j, k = 1, 2$$

where $\alpha_1 + \alpha_2 = 0$, $\beta_1 + \beta_2 = 0$, and $\gamma_{11} + \gamma_{12} = \gamma_{21} + \gamma_{22} = \gamma_{11} + \gamma_{21} = 0$ and e_{ijk} are independent Normal$(0,\sigma^2)$ random variables. Given the data

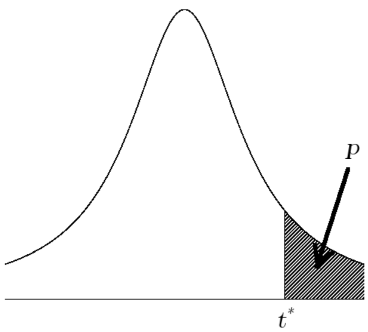
	y_{ijk}	i = 1	2
j	1	6,4	1,3
	2	7,5	1,5

test the hypothesis $H_0 : \alpha_1 = \alpha_2 = 0$ at level 0.05.

Problem 5. (20 points) Consider a random effects model

$$y_{ij} = \mu + a_i + e_{ij}, i = 1, \ldots, a, j = 1, \ldots, r$$

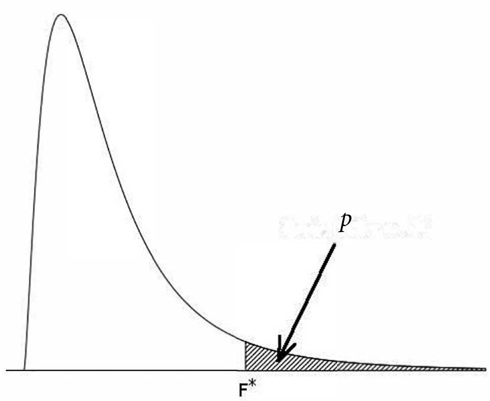
where the e_{ij}'s are independent Normal random variables with mean 0 and variance σ_e^2, the a_i's are independent Normal random variables with mean 0 and variance σ_a^2, and μ is non-random. Also, all e_{ij}'s and a_i's are mutually independent. Compute the following quantities.
(a - 4 pts) E$[y_{ij}]$
(b - 4 pts) Var$[y_{ij}]$
(c - 4 pts) Cov$[y_{ij}, y_{ij'}]$ for $j \neq j'$
(d - 4 pts) Cov$[y_{ij}, y_{i'j}]$ for $i \neq i'$
(e - 4 pts) Cov$[y_{ij}, y_{i'j'}]$ for $i \neq i'$ and $j \neq j'$



The critical value t^* is the value such that the area under the density curve of a t distribution with df degrees of freedom to the right of t^* is equal to p. It is also the value such that the area under the curve between $-t^*$ and t^* is equal to C.

t distribution critical values

	Upper-tail probability p				
df	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.25
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
	80%	90%	95%	98%	99%
	Confidence level C				



The critical value F^* is the value such that the area under the density curve of an F distribution with $df1$ degrees of freedom in the numerator and $df2$ degrees of freedom in the denominator to the right of F^* is equal to p.

F distribution critical values

$p = .05$

$df2$ \ $df1$	1	2	3	4	5
1	161.45	199.5	215.71	224.58	230.16
2	18.51	19.00	19.16	19.25	19.30
3	10.13	9.55	9.28	9.12	9.01
4	7.71	6.94	6.59	6.39	6.26
5	6.61	5.79	5.41	5.19	5.05
6	5.99	5.14	4.76	4.53	4.39
7	5.59	4.74	4.35	4.12	3.97
8	5.32	4.46	4.07	3.84	3.69
9	5.12	4.26	3.86	3.63	3.48
10	4.96	4.10	3.71	3.48	3.33
11	4.84	3.98	3.59	3.36	3.20
12	4.75	3.89	3.49	3.26	3.11
13	4.67	3.81	3.41	3.18	3.03
14	4.60	3.74	3.34	3.11	2.96
15	4.54	3.68	3.29	3.06	2.90
16	4.49	3.63	3.24	3.01	2.85
17	4.45	3.59	3.20	2.96	2.81
18	4.41	3.55	3.16	2.93	2.74
19	4.38	3.52	3.13	2.90	2.74
20	4.35	3.49	3.10	2.87	2.71

$p = .025$

$df2$ \ $df1$	1	2	3	4	5
1	647.79	799.50	864.16	899.58	921.85
2	38.51	39.00	39.17	39.25	39.30
3	17.44	16.04	15.44	15.10	14.88
4	12.22	10.65	9.98	9.60	9.36
5	10.01	8.43	7.76	7.39	7.15
6	8.81	7.26	6.60	6.23	5.99
7	8.07	6.54	5.89	5.52	5.29
8	7.57	6.06	5.42	5.05	4.82
9	7.21	5.71	5.08	4.72	4.48
10	6.94	5.46	4.83	4.47	4.24
11	6.72	5.26	4.63	4.28	4.04
12	6.55	5.10	4.47	4.12	3.89
13	6.41	4.97	4.35	4.00	3.77
14	6.30	4.86	4.24	3.89	3.66
15	6.20	4.77	4.15	3.80	3.58
16	6.12	4.69	4.08	3.73	3.50
17	6.04	4.62	4.01	3.66	3.44
18	5.98	4.56	3.95	3.61	3.38
19	5.92	4.51	3.90	3.56	3.33
20	5.87	4.46	3.86	3.51	3.29