Preliminary Examination - Multivariate Models

Thursday, August 18, 2022, 1:00-3:30 PM.

Answer all questions and show all work unless state otherwise.

F-Table will be provided.

1. Consider a paper manufacturer who is interested in three different pulp preparation methods (the methods differ in the amount of hardwood in the pulp mixture) and four different cooking temperatures for the pulp and who wishes to study the effect of these two factors on the tensile strength of the paper. Each replicate of a factorial experiment requires 12 observations, and the experimenter has decided to run three replicates. However, the pilot plant is only capable of making 12 runs per day, so the experimenter decides to run one replicate on each of the three days and to consider the days or replicates as blocks. On each of the three days, he conducts the experiment as follows. A batch of pulp is produced by one of the three methods under study. Then this batch is divided into four samples, and each sample is cooked at one of the four temperatures. Then a second batch of pulp is made using another of the three methods. This second batch is also divided into four samples that are tested at the four temperatures. This process is then repeated, using a batch of pulp produced by the third method. Assume that replicates or blocks are random factor while pulp preparation methods and temperatures are fixed factors). The data are shown in the table below.

	Replicate1			Replicate2			Replicate3		
Pulp Preparation Method	1	2	3	1	2	3	1	2	3
Temperature (°F)									
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

   (a) State the name of the statistical design used and state the features of the design such as the treatments and the experimental units.

   (b) Give an appropriate statistical model for the response variable and state the model assumptions.

   (c) Suppose that the following sum of squares (SS) were calculated using the data:
   SS(Replicates)=77.55, SS(Method)= 128.39, SS(Temperature)=434.08,
   SS(Replicates*Method)=36.28, SS(Method*Temperature)=75.17, SS(Total)=822.97.
   Write down the complete ANOVA table showing the columns: Source, SS, DF, MS, EMS and F-value.

   (d) The manufacturer wants to know whether there are significant differences in the mean tensile strengths of the paper for the different methods and the levels temperature. Analyze the data and draw conclusions at 5% significance level.

2. A company uses three methods(M) to measure a critical dimension of aircraft engine parts (P). 20 parts were chosen at random and each of the 20 parts was measured by each method

twice (two replicates). All measurements were made in a random order. The company is interested in comparing the methods and in estimating the variability of the dimension across parts.

The following sum of squares (SS) were obtained using the data. $SS(Total) = 1274, SS(M) = 2.617, SS(P) = 1185, SS(M*P) = 27$.

Use the notation y_{ijk} for the measurement from ith method, jth part, and kth replicate.

(a) Give a suitable model for y_{ijk} and state the assumptions.

(b) Find the correlation between y_{111} and y_{112}.

(c) Write down the ANOVA table that includes the columns: Source, DF, MS, EMS and F-value.

(d) Give the pair(s) of null and alternative hypotheses you would test to determine if there is a significant difference between the three methods. For each test, give the value of test statistic and the null distribution of the test statistic.

(e) Give an estimate of the correlation in part (a).

3. The following model was fit to the experimental data:

$$y_{ijkl} = \mu + \tau_i + b_j + c_{k(j)} + (\tau b)_{ij} + (\tau c)_{ik(j)} + e_{ijkl},$$

with $i = 1,2,3; j = 1,2,3,4; k = 1,2,3; l = 1,2,3,4,5$; where μ and τ_j are population parameters; and b_j, $c_{k(j)}, (\tau b)_{ij}, (\tau c)_{ik(j)}$ and e_{ijkl} are independent random variables with $N(0, \sigma_B^2)$, $N(0, \sigma_{C(B)}^2)$, $N(0, \sigma_{AB}^2)$, $N(0, \sigma_{AC(B)}^2)$, and $N(0, \sigma_e^2)$ distributions, respectively.

Source	DF	MS	EMS
A		24.5	$\sigma_e^2 + r\sigma_{AC(B)}^2 + cr\sigma_{AB}^2 + Q_A$
B		19.7	$\sigma_e^2 + r\sigma_{AC(B)}^2 + ar\sigma_{C(B)}^2 + cr\sigma_{AB}^2 + acr\sigma_B^2$
AB		8.9	$\sigma_e^2 + r\sigma_{AC(B)}^2 + cr\sigma_{AB}^2$
C(B)		7.5	$\sigma_e^2 + r\sigma_{AC(B)}^2 + ar\sigma_{C(B)}^2$
AC(B)		6.8	$\sigma_e^2 + r\sigma_{AC(B)}^2$
Error		5.8	σ_e^2

(a) Complete the above ANOVA table for the experiment by filling in the degrees of freedom.

(b) Test for a significant AB interaction at 5% significance level.

(c) Test for a significant B main effect at 5% significance level.

(d) Compute the variance of the difference in treatment means for levels 1 and 2 of Factor A: $\bar{y}_{1...} - \bar{y}_{2...}$. Provide an estimate of this variance and the associated degrees of freedom.

4. Let a $\boldsymbol{X} = (X_1, ..., X_p)'$ be a p-dimensional random vector of p scalar random components $X_1, .., X_p$. Assume $\boldsymbol{X}$ has a multivariate normal distribution with unknown mean $\boldsymbol{\mu}$ and unknown variance-covariance matrix Σ. Suppose a random sample of size n was obtained from the distribution of $\boldsymbol{X}$, and let $\bar{X}$ be the sample mean and S be the sample variance-covariance matrix.

You may give your answers in terms of percentiles of known distributions, e.g., $t_{0.05}(df = 10)$.

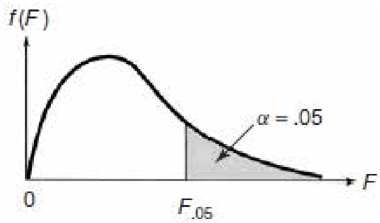
(a) Give the maximum likelihood estimates of $\boldsymbol{\mu}$ and Σ (no need to show work for this part).

(b) Gove a 95% simultaneous confidence region for $\boldsymbol{\mu}$ (no need to show work for this part).

(c) Suppose we want to test the null hypothesis that the the scalar random variables $X_1, .., X_n$ are independent against the alternative that they are not. State the null hypothesis H_0 and the alternative hypotheses H_1 in terms of the parameters of the multivariate distribution and find the likelihood ratio test for testing the hypothesis H_0 vesrus H_1. Show work for this part.

Give an approximate rejection region with significance level α.

Table D.4 Critical values for the F Statistic: $F_{.05}$



ν_2 \ ν_1	Numerator Degrees of Freedom								
DENOMINATOR DEGREES OF FREEDOM	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88