

ISyE6413
 First Midterm Examination October 6, 2004
 (Total : 50 points)

Name :

Letter Grade () or Pass-Fail ()?

Problem 1 (12 pts)

Questions (a)-(d) involve the analysis of some data on cigarette consumption. The data was collected in 1970. The variable **perfem** is the percentage of females living in a state. Data are collected on each of the 50 states of the USA and Washington DC. The purpose of the analysis is to see how cigarette consumption is related to the other variables.

- (a) (3 pts) A simple linear regression model was fitted trying to predict the cigarette consumption using the percentage of females and the following partial output was obtained:

Coefficients:

	Value	Std. Error	<i>t</i> Value	$Pr(> t) = p - value$
(Intercept)	-93.4260	207.8126	-0.4496	0.6550
perfem	4.2191	4.0777	1.0347	0.3059

Residual standard error: 32.05 on 49 degrees of freedom

Multiple R-Squared: 0.02138

The slope is not statistically significant but the cigarette consumption of males and females is known to differ. Offer the best explanation for this apparent contradiction.

perfem does not vary much from state to state - collinear with the intercept term.

- (b) (3 pts) Suppose we replaced the percentage of females by the percentage of males in the above regression model. Give the regression coefficients for this new model. Would the standard error for the slope change?

$y = -93.426 + 4.2191 \times (100 - x) = 328.484 - 4.2191x$. *se will not change.*

- (c) (2 pts) In a regression problem, what (if any) is the unit of the *t*-statistics?

no unit.

- (d) (4 pts) When we look at the ANOVA table for a regression model, it is possible that the *p*-value reported by MINITAB is 0.000, which does not mean that it is actually zero, just extremely small. Would it be possible (for some other data) to have a *p*-value of exactly zero in an *F*-statistics in a regression? Explain clearly.

For p-value to be zero, the F-statistic would need to be infinite. This can only happen if RSS of the model is zero - in other words, if the data fit the model perfectly.

Let n = number of observations, and p = number of parameters in the regression model. Someone claims that setting $n = p$ will give a p-value of exactly zero. Is the claim correct ?

setting $n=p$ does not work since this one won't have any degree of freedom for error and hence, no test is possible!

Problem 2 (10 pts)

An experiment is being performed to determine the optimum temperature and conveyor speed that will yield a desired level of hardness in copper after annealing (a heat treatment operation). Three levels of temperature viz. 1000, 1050 and 1100 degree F and three levels of conveyor motor rpm viz. 1400, 1600 and 1800 are chosen for the experiment. The experimenter has only 9 test pieces or experimental units, three each from three suppliers A, B and C. It is suspected that there may be a supplier-to-supplier variation with respect to the hardness of raw material. Further, three furnaces F1, F2 and F3 are available for the experiment, and the department head wants all of them to be used for conducting the experiment so that the total time required is less. The experimenter feels the three furnaces are not identical in behavior.

- (a) (2 pt) What kind of design should be used for conducting this experiment?

Since there are two experimental factors (temperature, conveyor speed) and two block factors (supplier, furnace) a Graeco Latin-Square has to be used.

- (b) (8 pts) Design an appropriate experiment and write down the planning matrix.

Note: Planning matrix means the physical layout of the experiment, i.e, the experimental plan that you would hand over to the operators so that they can conduct the trials for you without any ambiguity.

Problem 3 (10 pts)

Consider an experiment with four factors A, B, C and D . Among them, A and B have two levels each, C has three and D has four levels. Two replications are conducted with all possible combinations.

- (a) (2 pts) In the analysis of experimental data with a four-way layout, what will be the total degrees of freedom?

$$2^2 \times 3 \times 4 \times 2 - 1 = 95.$$

- (b) (4 pts) In the ANOVA table, how many degrees of freedom will you have corresponding to error?

We have 7 df corresponding to 4 main effects, 17 df corresponding to $\binom{4}{2} = 6$ two-factor interactions, 17 df corresponding to $\binom{4}{3} = 4$ three-factor interactions and 6 df corresponding to one four-factor interaction. Thus residual df = $95 - (7+17+17+6) = 48$. Alternatively, residual df = $IJKL(N - 1) = 2 \times 2 \times 3 \times 4 \times (2-1) = 48$

- (c) (4 pts) Suppose the F-statistic for testing the significance of interaction $B \times D$ is computed as 3.16. Will you declare $B \times D$ significant at 5% level of significance? Explain.

Since computed $F > F_{3,48,0.05} \approx 2.80$, $B \times D$ will be declared significant at 5% level.

Problem 4 (4 pts)

Recall that a BIBD has t treatments, and b blocks of size k . Each treatment is replicated r times, such that each pair of treatments appear in the *same* number (denoted by λ) of blocks. For the following two cases, explain whether it is possible to construct a BIBD or not.

- (a) $t = 8, k = 4, b = 6, r = 3$ and $\lambda = 2$.

$$r(k-1) \neq \lambda(t-1).$$

- (b) $t = 4, k = 4, b = 3, r = 3$ and $\lambda = 3$.

$$t \not\geq k.$$

Problem 5 (10 pts)

Questions (a)-(c) involve the analysis of some data on length of cuckoo eggs. A study by E.B. Chance in 1940 called "The Truth About the Cuckoo" demonstrated that cuckoos return year after year to the same territory and lay their eggs in the nests of other (host) birds. The eggs are then adopted and hatched by the host birds. It is of interest to see whether the length of the cuckoo eggs vary with the type of host bird that hatches the eggs.

From a large population of host birds, six were chosen at random. The lengths of 120 cuckoo eggs found in the nests of these six birds were measured. The type of birds and the number of eggs measured in their nests is given below:

1. Meadow Pipit (MDWP) - 45 eggs were found and measured.
2. Tree Pipit (TRP) - 15 eggs were found and measured.
3. Hedge Sparrow (HGSPRW) - 14 eggs were found and measured.
4. Robin(RBN) - 16 eggs were found and measured.
5. Pied Wagtail (PWGTL) - 15 eggs were found and measured.
6. WREN (WREN) - 15 eggs were found and measured.

- (a) (6 pts) Fill in the gaps in the above ANOVA table.
- (b) (1 pt) Is there is significant difference among host birds with respect to length of cuckoo eggs?

Yes.

Table 1: ANOVA Table, cuckoo eggs' length data

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F statistic
Host birds	-	42.940	88.588	-
Residual total	-	-	-	-
	-	137.188		

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F statistic
Host birds	5	42.940	8.588	10.39
Residual	114	94.248	0.827	
total	119	137.188		

(c) (3 pts) Obtain an estimate of the host bird to host bird variance with respect to length of eggs.

Here, $k = 6, n_1 = 45, n_2 = 15, n_3 = 14, n_4 = 16, n_5 = 15, n_6 = 15$.

Hence, $n' = \frac{1}{k-1} \left[\sum_{i=1}^k n_i - \frac{\sum_{i=1}^k n_i^2}{\sum_{i=1}^k n_i} \right] \approx 19$.

Thus, $\hat{\sigma}_\tau^2 = \frac{MST - MSE}{n'} = 0.4085$.