Homework I (2023)

Please solve the following problems and then submit the pdf copy of them.

1. In class, we have learned the sample size calculation for two samples. That is, $y_{i1}, \ldots, y_{in} \overset{iid}{\sim} N(\mu_i, \sigma_j^2)$ for $i = 1, 2$. Show that under $H_0 : \Delta = \Delta_0$ versus $H_1 : \Delta \neq \Delta_0$,
$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2},$$
where $\Delta = \mu_1 - \mu_2$ is the treu difference in means.

2. We want to compare fasting serum cholesterol levels in subjects over 21 years of age belonging to two ethnic groups. A random sample of 20 subjects from group 1 had a mean level of 180 mg/dL with a standard deviation of 41 mg/dL. The corresponding values for a random sample of 30 subjects were 165 mg/dL and 36 mg/dL.

   (a) State the assumptions under which we are justified in using the independent sample $t-$test procedures for comparing mean cholesterol levels in the two groups.

   (b) Perform a hypothesis test to see if the true mean cholesterol level in group 1 is at least 5 mg/dL higher than that in group 2.

   (c) A subject presents with a measured cholesterol level of 210 mg/dL. On the basis of the given information, would it be reasonable to assume that the subject does not belong to group 1? Explain.

   (d) Construct a 95% confidence interval for the differential true mean cholesterol levels in the two groups; once assuming the population variances to be equal, then again without the equal variance assumption.

   (e) Perform a test to see if, on the basis of the given data, there is reason to conclude that the variances of the two populations are unequal. Write your conclusion.

3. The human resources department of a large manufacturing company is interested in comparing four employee training programs: computer assisted training on site (CON), computer assisted training off site (COF), video tape training on site (VON), and video tape training off site (VOF). Data below are scores that measure weekly average increase in productivity for 6 employees trained under each of the four training programs.

Program	Scores					
CON	16	18	19	21	24	20
COF	10	13	10	8	12	13
VON	21	23	19	26	22	24
VOF	12	16	13	14	16	13

(a) Using appropriate notation, write an ANOVA model for the data. Explain all terms in your model.

(b) Assuming the model in part (a), construct the appropriate ANOVA table. Perform an hypothesis test to see if the mean productivity score differ for the four mehtods.

(c) Construct a 95% confidence interval for the differential mean productivity scores for CON and COF. On the basis of the interval, which method would you recommend? Why?

(d) Construct three mutually orthogonal contrasts θ_1, θ_2, θ_3 such that

- θ_1 can be used to compare the computer and video methods.
- θ_2 can be used to compare the on site and off site methods.
- θ_3 can be used to compare the difference between computer and video methods when they are used on site with the corresponding difference when they are used off site.

(e) Test significance of the three contrasts in (d) using an experimentwise error of 5%. Interpret the results of your tests.

(f) Perform a pairwise multiple comparison of the four treatment means using the Tukey method.

4. In the single-factor ANOVA model, $y_{ij} = \mu+\tau_i+\epsilon_{ij}$ for $i = 1, \cdots, a$ and $j = 1, \cdots, n$ where $\sum_i \tau_i = 0$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. Then prove the following results:

(a) $E(MS_{Treatment}) = \sigma^2 + \frac{n}{a-1}\sum_{i=1}^{a} \tau_i^2$.

(b) $E(MS_E) = \sigma^2$.

(c) Under $H_0 : \tau_1 = \cdots = \tau_a = 0$, prove that

$$F = MS_{Treatment}/MS_E \sim F_{a-1,a(n-1)}.$$