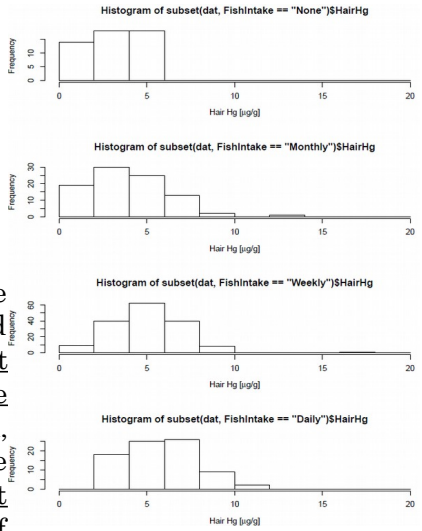
Evidence based healthcare special lecture I Final mini-test 19th July 2017, 18:30-20:00, D101
(Please see opposite side of the paper)

(Example of the answer)

1. Please specify the wrong points in explanation and/or method to analyze and suggest how to improve it (if no wrong point, answer so) for each issues underlined below.

(1) In R district with population size of about 38,000 located along the sea in a developing country, the civil registration system has been established since several decades ago. When we see there with Google Earth, almost equal-sized 50 villages are scattered. Recently the people with high fish and whale intake are reported to suffer from neural damage symptoms. Due to the possibility of mercury poisoning, (A) randomly selected 1% sample of civil registration, 380 residents were interviewed about the frequency of fish/whale intake, such as [1] none or rarely, [2] monthly, [3] weekly, [4] daily, and the hair mercury concentrations of them were measured. The results (of which raw data are available from http://minato.sip21c.org/fish-Hg-2017.txt; variables are PID as personal ID number, HairHg as hair mercury concentration, HighHairHg is 1 if HairHg>=5, otherwise 0, and FishIL is one of 4 categories shown above) were summarized below.

Eating fish or whale frequency	N	Median (Hg μg/g hair)	Mean ± SD (Hg μg/g hair)	High Hg ≥5 μg/g hair
1. None	50	3.31	3.14 ± 1.74	8
2. Monthly	90	3.84	3.92 ± 2.21	29
3. Weekly	160	4.97	5.04 ± 2.15	80
4. Daily	80	5.82	5.73 ± 1.94	52



There are two approaches to analyze the relationships between fish/whale intake and hair mercury. First, the independence between High Hg and eating fish/whale frequency can be analyzed. (B) Fisher's exact test resulted in p<0.001 and the null hypothesis was rejected, so that the relationship between the two variables is statistically significant. Second, the effect of fish/whale intake on hair mercury concentration can be analyzed. (C) Welch's one-way ANOVA resulted in F-value of 25.591, first d.f. of 3, second d.f. of 163.13, and p<0.001. Then pairwise comparisons of hair Hg levels between all pairs among 4 fish/whale intake frequencies can be conducted by repeated use of Welch's t-test.

(A) It's OK, but it is too time-consuming to visit randomly sampled 380 out of 38,000 individuals. Thus the cluster sampling (randomly sample 3 to 5 villages out of 50, where all residents will be recruited as subjects) is better option.

(B) Correct.

(C) Repeated use of Welch's t-test increases type I error. Adjustment of p-values for multiple comparisons is needed, using FDR or Holm's method, otherwise, Tukey's HSD is recommended to use.

(2) In Japan, to evaluate the effect of long-term intake of the food A (the reduction of blood pressure is expected), an intervention experiment was conducted for 5 hypertension patients, where the only intervention was daily intake of food A for 6 months. The changes of systolic blood pressure (mmHg) between the 2 timings (before intervention and after 6 months) were 160→145, 150→125, 170→155, 155→135, 145→130. The result of paired t-test was p=0.0008, so that the food A is proved to have a significant blood pressure reducing effects.

Japan has the large seasonal changes of climate and foods, which may affect the result. Thus control group is needed. Randomly assign patients into 2 groups (taking A and not forced to take A), follow up them for 6 months, then compare the changes of SBP (possibly DBP too) between 2 groups or conduct the repeated measures ANOVA including not only before/after, but also in-between data.

(3) The gold standard method A can measure the concentration of biochemical marker for disease X, where the concentration exceeds a specific threshold value. A cheaper and more rapid new method B was developed. Validity of B can be confirmed by showing the fact that the difference between the measurements by A and B for the sufficient number of X patients and healthy volunteers is not statistically significant by paired t-test.

"Difference between the measurements by A and B is not significant" is not enough. High correlation between two series and the difference unaffected by absolute level are also needed. Thus Bland-Altman plot should be conducted. Then, the substantial equality of AUCs of ROC curves between A and B (or AUC by B is larger than by A) should be confirmed.

(4) The 44 chronic hepatitis patients were randomly divided into 2 groups. Treatment group was treated by prednisolone, the other (control group) was just observed. At the end of the study, 11 patients lived in treatment group, 6 lived in control group, but the result of Fisher's exact test was not significant (p=0.215). The months until death or censoring (lived at the end of the study) were recorded in http://minato.sip21c.org/hepatitis-2017.txt as **time**, with **flag** (1 if died, 0 if still lived) and **group** (1 for treatment group, 2 for control group). Only for the patients who died during the study, the mean survival months (80 months in 11 treatment group and 31.5 months in 16 control group) were compared by Welch's t-test, then p=0.02, so that prednisolone has statistically significant effect of lengthen survival for chronic hepatitis patients.

If the data of died patients are exclusively used, the data of relatively longer survived patients are not used, and thus the survival time is underestimated. When Kaplan-Meyer method is applied to the all patients data, median survival times are 146 months and 40.5 months for the treatment group and control group, respectively. Log-rank test results in p=0.0309, so that the use of prednisolone significantly prolong the survival time (In this data, the final conclusion is same, but the methodology is not adequate).

2. Please explain the prevalence proportion as a disease amount in a population. Explanation for the study design needed to calculate prevalence proportion has to be included.

   Prevalence proportion is obtained when the cross-sectional study is applied, which is the proportion of disease patients among total subjects. Meaning is the amount of disease burden on the population.

3. Glucose tolerance test was conducted for 10 healthy volunteers, the blood inorganic phosphate concentration was measured at 6 timings (before glucose intake, just after the glucose intake, 30 min later, 1 hour later, 2 hours later and 3 hours later). Please explain what kind of statistical method to test the change of measurement with time is applicable.
   Friedman's test is applicable. However, if there is no extreme outlier, repeated measures ANOVA is better to use.

4. Whether the ability of flash memory is improved by tea drinking or not was investigated for 10 healthy volunteers. The result is shown below. Please test whether tea drinking improves flash memory or not. P-value is needed. You can use computer software or calculator, but manual calculation is possible if you use 97.5% point of t-distribution with d.f. 9 is 2.262 and either of $\sqrt{2}$=1.414, $\sqrt{3}$=1.732, or $\sqrt{5}$=2.236.

Scores before drinking tea	7	8	7	9	3	7	5	8	7	6
Scores after drinking tea	9	9	7	10	4	10	6	9	8	10

   Execute the R code t.test(c(7, 8, 7, 9, 3, 7, 5, 8, 7, 6), c(9, 9, 7, 10, 4, 10, 6, 9, 8, 10), paired=TRUE), then you get p=0.003. The null-hypothesis that the score does not change is rejected. Consequently, after drinking tea, the short memory is significantly improved.

5. The RCT (Randomized Controlled Trial) to test new drug for abdominal pain relief intended to check the superiority of the new drug over the conventional drug. Based on previous studies, the pain-killer effect of the conventional drug was found in 75% people. If more than 85% people report the pain reduction by taking the new drug, the new drug has clinical importance. Please calculate needed sample size for this RCT for 2-tailed chi-square test with 5% significance level and 80% power, and assuming the same size of 2 groups.

   The result of power.prop.test(p1=0.75, p2=0.85, sig.level=0.05, power=0.8) is **250** each.