Class 19 in-class problems, 18.05, Spring 2022

Concept questions

Concept question 1. t-test odds

We run a two-sample t-test for equal means, with $\alpha = 0.05$, and obtain a p-value of 0.04. What are the odds that the two samples are drawn from distributions with the same mean?

(a) 19/1   (b) 1/19   (c) 1/20   (d) 1/24   (e) unknown

Concept question 2. Multiple testing

(a) Suppose we have 6 treatments and want to know if the average recovery time is the same for all of them. If we compare two at a time, how many two-sample t-tests do we need to run?

(i) 1   **(ii)** 2   **(iii)** 6   **(iv)** 15   **(v)** 30

(b) Suppose we use the significance level 0.05 for each of the 15 tests. Assuming the null hypothesis, what is the best estimate of the probability that we reject at least one of the 15 null hypotheses?

(i) < 0.05 **(ii)** 0.05   **(iii)** 0.10   **(iv)** > 0.25

Board questions

Problem 1. Khan's restaurant

Sal is thinking of buying a restaurant and asks about the distribution of lunch customers. The owner provides row one below. Sal records the data in row two himself one week.

	M	T	W	R	F	S
Owner's distribution	0.1	0.1	0.15	0.2	0.3	0.15
Observed # of cust.	30	14	34	45	57	20

Set the significance level ahead of time.

H_0: the owner's distribution is correct.

H_A: the owner's distribution is not correct.

Compute both G and X^2.

Run a chi-square goodness-of-fit test on the null hypotheses:

Problem 2. Genetic linkage

In 1905, William Bateson, Edith Saunders, and Reginald Punnett were examining flower color and pollen shape in sweet pea plants by performing crosses similar to those carried out by Gregor Mendel.

The genes for color and shape are given by:

Purple flowers (P) is dominant over red flowers (p).
Long seeds (L) is dominant over round seeds (l).

In the first generation there were only two genetic types PPLL and ppll. There initial cross was always PPLL with ppll. So this always resulted in PpLl in the second generation. The second generation plants were then crossed randomly with each other to make the third generation.

F0: PPLL x ppll (initial cross)
F1: PpLl x PpLl (all second generation plants were PpLl)
F2: 2132 plants (third generation)

H_0 = independent assortment: color and shape are inherited independently.

Here is the data from their experiment.

	purple, long	purple, round	red, long	red, round
Expected	?	?	?	?
Observed	1528	106	117	381

Determine the expected counts for F_2 under H_0 and find the p-value for a Pearson chi-square test. Explain your findings biologically.

Problem 3. Recovery
The table shows recovery time in days for three medical treatments.

(a) Set up and run an F-test testing if the average recovery time is the same for all three treatments. Use significance level 0.05.

(b) Based on the test, what might you conclude about the treatments?

T_1	T_2	T_3
6	8	13
8	12	9
4	9	11
5	11	8
3	6	7
4	8	12

Note: For $\alpha = 0.05$, the critical value of $F_{2,15}$ is 3.68.

Problem 4. Chi-square for independence
(From Rice, *Mathematical Statistics and Data Analysis*, 2nd ed. p.489)

Consider the following contingency table of counts

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

Use a chi-square test with significance level 0.01 to test the hypothesis that the number of marriages and education level are independent.

Question not used in class: z-test
We have 16 independent sample values $x_1, \ldots, x_{16}$ drawn from a Normal$(\theta, 8^2)$ distribution.

Suppose the sample mean $\bar{x} = 4$. Run a z-test on this data for the null hypothesis $\theta = 2$ vs the alternative $\theta \neq 2$. Choose a significance of $\alpha = 0.04$.