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

Solution: (e) unknown. Frequentist methods only give probabilities for data under an assumed hypothesis. They do not give probabilities or odds for hypotheses. So we don't know the odds for distribution means

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

Solution: **(a)** (iv) 6 choose 2 = 15.

(b) (iv) Greater than 0.25.

Under H_0 the probability of rejecting for any given pair is 0.05. Because the tests aren't independent, i.e. if the group1-group2 and group2-group3 comparisons fail to reject H_0, then the probability increases that the group1-group3 comparison will also fail to reject.

We can say that the following 3 comparisons: group1-group2, group3-group4, group5-group6 are independent. The number of rejections among these three follows a $\text{binom}(3, 0.05)$ distribution. The probability the number is greater than 0 is $1 - (0.95)^3 \approx 0.14$.

Even though the other pairwise tests are not independent, they do increase the probability of rejection. In simulations of this with normal data, the false rejection rate was about 0.36.

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:

Solution: The total number of observed customers is 200. The table of expected (under H_0) and observed counts is

	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
Expected # of cust.	20	20	30	40	60	30

So,

$$G = 2\sum O_i \log(O_i/E_i) = 11.39$$

$$X^2 = \sum \frac{(O_i - E_i)^2|}{E_i} = 11.44$$

$df = 6 - 1 = 5$ (6 cells, compute 1 value –the total count– from the data)

$$p = \texttt{1-pchisq(11.39,5) = 0.044.}$$

So, at a significance level of 0.05 we reject the null hypothesis in favor of the alternative that the owner's distribution is wrong.

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.

Solution: For color, all F1 generation flowers have genotype Pp.

So, we expect F2 to split 1/4, 1/2, 1/4 between PP, Pp, pp. So, for the phenotype, we expect F2 to have 3/4 purple and 1/4 red flowers.

Similarly for LL, Ll, ll: we expect F2 to have 3/4 long and 1/4 round seeds.

Assuming H_0, color and shape are independent. So, we can multiply probabilities to get the following probabilitiy table for phenotypes in F2

	Long	Round	
Purple	9/16	3/16	3/4
Red	3/16	1/16	1/4
	3/4	1/4	1

We have a total of 2132 plants in F2, so we expect $2132 \times 9/16 \approx 1199$ purple color-long seed flowers. Likewise for the other phenotypes. The table of expected counts is then:

	purple, long	purple, round	red, long	red, round
Expected	1199	400	400	133
Observed	1528	106	117	381

Using R we compute: $G = 972.0$, $X^2 = 966.6$.

The degrees of freedom = 3 (4 cells - 1 cell needed to make the total work out).

The p-values for both statistics are effectively 0.

At almost all significance levels we would reject H_0 in favor of the alternative that the genes are not indpendent.

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.

Solution: The null hypothesis H_0 is that the means of the 3 treatments are the same. H_A is that they are not.

We will run an F-test (ANOVA). Our test statistic f is computed following the procedure given in the slides and notes

We have $n = 3$ groups of data with $m = 6$ data points each.

MS_B = between group variance $= \dfrac{m}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2 = 42$

MS_W = within group variance = average of the 3 sample standard deviations = 68/15.

$$\text{test statistic: } f = \frac{\text{MS}_B}{\text{MS}_W} \approx 9.26$$

Under H_0: $f \sim F_{n-1,\, n(m-1)} = F_{2,15}$.

So, the p-value

$$p = P(F > f|H_0) = \texttt{1 - pf(9.26,2,15)} \approx 0.0024.$$

So, we reject H_0 in favor of the hypothesis that the means of three treatments are not the same.

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.

Solution: The null hypothesis is that the cell probabilities are the product of the marginal probabilities. Assuming the null hypothesis we estimate the marginal probabilities in orange and multiply them to get the cell probabilities in blue.

Education	Married once	Married multiple times	Total
College	0.365	0.061	611/1436
No college	0.492	0.082	825/1436
Total	1231/1436	205/1436	1

We then get expected counts by multiplying the cell probabilities by the total number of women surveyed (1436). The table shows the observed, expected counts:

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

We then have

$$G = 16.55 \qquad \text{and} \qquad X^2 = 16.01$$

The number of degrees of freedom is $(2-1)(2-1) = 1$. (We can count this: we needed the marginal counts to compute the expected counts. Now setting any one of the cell counts determines all the rest because they need to be consistent with the marginal counts from the data.) So, we get

$$p = \texttt{1-pchisq(16.55,1) = } 0.000047$$

Therefore we reject the null hypothesis in favor of the alternate hypothesis that number of marriages and education level are not independent

Question not used in class: z-test

We have 16 independent sample values $x_1, \dots, x_{16}$ 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 $\alpha = 0.04$.

Solution: The z-statistic is $z = \dfrac{\bar{x} - 2}{8/\sqrt{16}} = 1$.

This is a two-sided test and $z > 0$, so $p = 2P(Z > 1) \approx 0.32$.

Since $p > \alpha$, the data does not support rejecting H_0.