PH 1700 Homework 1

Xu Zuo, M.S.

Microbiology

A study was conducted to demonstrate that soy beans inoculated with nitrogen-fixing bacteria yield more and grow adequately without expensive environmentally deleterious synthesized fertilizers. The trial was conducted under controlled conditions with uniform amounts of soil. The initial hypothesis was that inoculated plants would outperform their uninoculated counterparts. This assumption is based on the facts that plants need nitrogen to manufacture vital proteins and amino acids and that nitrogen-fixing bacteria would make more of this substance available to plants, increasing their size and yield. There were 8 inoculated plants (*I*) and 8 uninoculated plants (*U*). The plant yield as measured by pod weight for each plant is given in Table 2.20.

2.35 Compute appropriate descriptive statistics for *I* and *U* plants.

TABLE 2.20 Pod weight (g) from inoculated (*I*) and uninoculated (*U*) plants

Sample number	*I*	*U*
1	1.76	0.49
2	1.45	0.85
3	1.03	1.00
4	1.53	1.54
5	2.34	1.01
6	1.96	0.75
7	1.79	2.11
8	1.21	0.92

Note: The data for this problem were supplied by David Rosner.

2.36 Use graphic methods to compare the two groups.

2.37 What is your overall impression concerning the pod weight in the two groups?

Answer 2.35

Table 1 is a summary table of I and U plants. Each group includes 8 samples. The table shows the descriptive statistics of pod weights from two groups, including mean, standard deviation, minimum and maximum.

Table 1. Descriptive statistics for I and U plants. Obs: number of observations; Std. Dev.: standard deviation; Min: minimum; Max: maximum.

Variable	Obs	Mean	Std. Dev.	Min	Max
I	8	1.63375	.4198958	1.03	2.34
U	8	1.08375	.5097881	.49	2.11

Answer 2.36

Figure 1 shows the box plots for the I and U groups. The distribution shows that plants in the I group generally have heavier pod weights. Sample 7 in the U group has the pot weight of 2.11 and was plotted as an outliner in the graph.

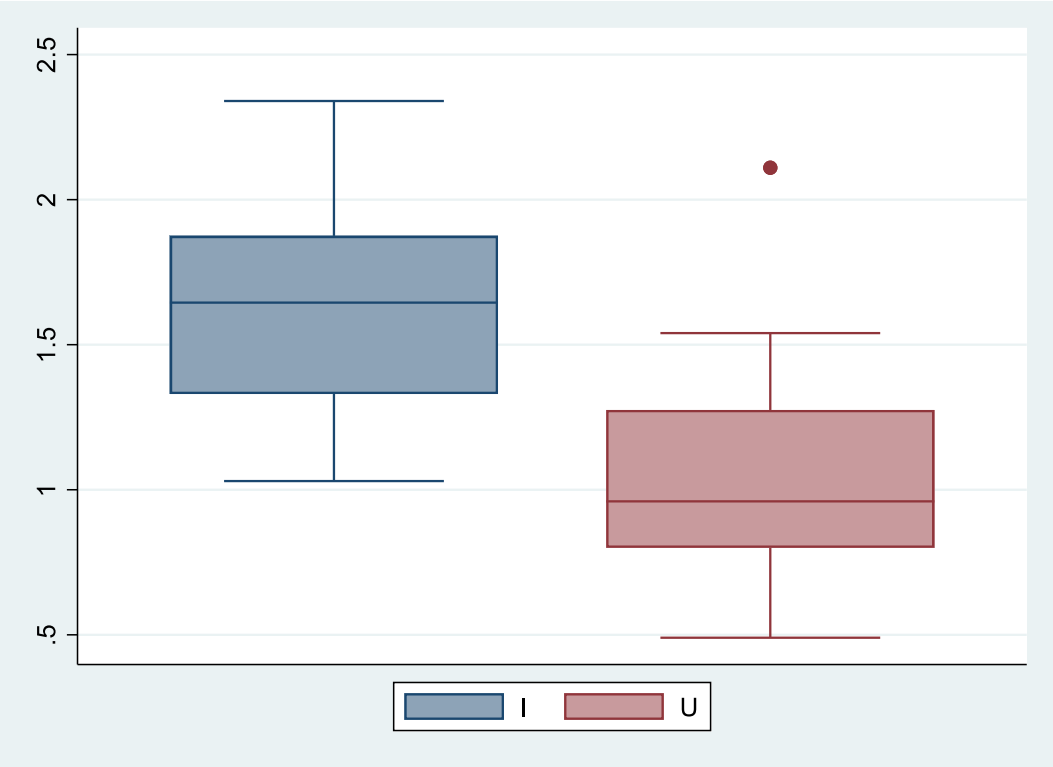

Figure 1. Box plots of I and U groups.

Answer 2.37

Plants in the I group generally grow better than plants in the U group, since the I group has higher mean and lower standard deviation in terms of pod weights. If the pod weight is one of the metrics that assesses the plant growth, then the statistics indicate that inoculation is effective for soybean growth.

Endocrinology

In Section 2.10, we described Data Set BONEDEN.DAT (at www.cengagebrain.com) concerning the effect of tobacco use on BMD.

2.38 For each pair of twins, compute the following for the lumbar spine:

A = BMD for the heavier-smoking twin − BMD for the lighter-smoking twin = $x_1 - x_2$
B = mean BMD for the twinship = $(x_1 + x_2)/2$
C = 100% × (A/B)

Derive appropriate descriptive statistics for C over the entire study population.

2.39 Suppose we group the twin pairs according to the difference in tobacco use expressed in 10 pack-year groups (0–9.9 pack-years/10–19.9 pack-years/20–29.9 pack-years/30–39.9 pack-years/40+ pack-years). Compute appropriate descriptive statistics, and provide a scatter plot for C grouped by the difference in tobacco use in pack-years.

2.40 What impression do you have of the relationship between BMD and tobacco use based on Problem 2.39?

Answer 2.38

Table 2. Descriptive statistics for variable C. Obs: number of observations; Std. Dev.: standard deviation; Min: minimum; Max: maximum.

Variable	Obs	Mean	Std. Dev.	Min	Max
C	41	-4.949668	12.48342	-38.09524	19.81982

Answer 2.39

Table 3. Descriptive statistics for variable C by tobacco use group. Cat_pyr: categories of 10 pack-year groups; N: number of observations; sd: standard deviation.

cat_pyr	N	mean	sd	min	max
0-9.9	7	1.950075	8.263643	-6.451613	16.57459
10-19.9	14	-2.18482	10.45423	-14.08451	16.77419
20-29.9	8	-10.17456	16.69339	-38.09524	11.92053
30-39.9	6	-8.300975	2.885653	-12.63158	-4.938272
40+	6	-9.132847	17.7739	-35.89743	19.81982
Total	41	-4.949668	12.48342	-38.09524	19.81982

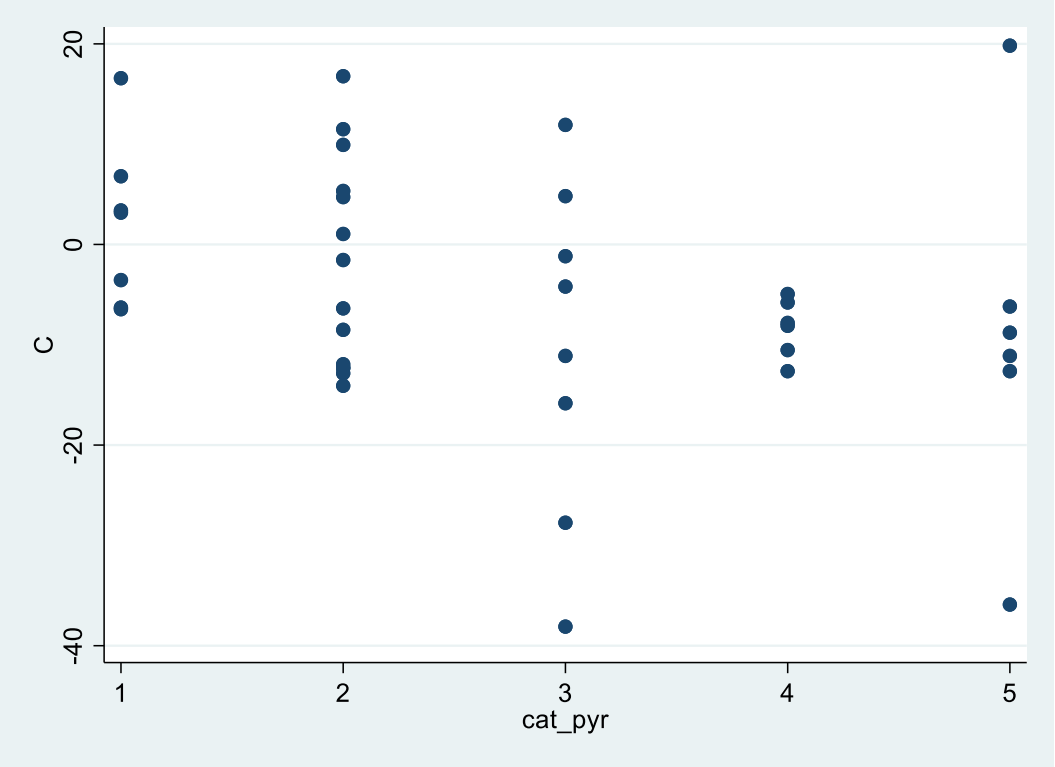

Figure 2. Scatter plot for variable C by tobacco use group. Cat_pyr: categories of 10 pack-year groups. Each group number corresponds to a pack year group: 1-> 0-9.9 pack years; 2-> 10-19.9 pack years; 3-> 20-29.9 pack years; 4-> 30-39.9 pack years; 5-> 40+ pack years.

Answer 2.40

The figure implies that C is decreasing as the difference in pack-years increases. This might indicate that the heavier-smoking twin has a smaller lumbar spine bone mineral density, which suggests that tobacco use has a negative relationship with lumbar spine bone mineral density.

Estimates of the prevalence of Alzheimer's disease have recently been provided by Pfeffer et al. [8]. The estimates are given in Table 3.5.

Suppose an unrelated 77-year-old man, 76-year-old woman, and 82-year-old woman are selected from a community.

3.18 What is the probability that at least one of the three people has Alzheimer's disease?

3.22 Suppose we know two of the three people have Alzheimer's disease. What is the conditional probability that they are both younger than 80 years of age?

TABLE 3.5 Prevalence of Alzheimer's disease (cases per 100 population)

Age group	Males	Females
65–69	1.6	0.0
70–74	0.0	2.2
75–79	4.9	2.3
80–84	8.6	7.8
85+	35.0	27.9

Answer 3.18

We denote A = {77-year-old man has Alzheimer's disease}, B = {76-year-old woman has Alzheimer's disease}, C = {82-year-old woman has Alzheimer's disease}.

$$\Pr(A \cup B \cup C) = 1 - \Pr(\bar{A} \cap \bar{B} \cap \bar{C}) = 1 - (1 - 0.49) \times (1 - 0.23) \times (1 - 0.78) = 0.143$$

Answer 3.22

$$Pr(both < 80\ years\ old | two\ have\ Alzheimer's\ disease) = \frac{0.49 \times 0.23 \times (1 - 0.78)}{0.49 \times 0.23 \times (1 - 0.78) + 0.49 \times (1 - 0.23) \times 0.78 + (1 - 0.49) \times 0.23 \times 0.78} = 0.16$$

The Chinese Mini-Mental Status Test (CMMS) consists of 114 items intended to identify people with Alzheimer's disease and senile dementia among people in China [14]. An extensive clinical evaluation of this instrument was performed, whereby participants were interviewed by psychiatrists and nurses and a definitive diagnosis of dementia was made. Table 3.13 shows the results obtained for the subgroup of people with at least some formal education.

Suppose a cutoff value of ≤ 20 on the test is used to identify people with dementia.

3.87 What is the sensitivity of the test?

3.88 What is the specificity of the test?

TABLE 3.13 Relationship of clinical dementia to outcome on the Chinese Mini-Mental Status Test

CMMS score	Nondemented	Demented
0–5	0	2
6–10	0	1
11–15	3	4
16–20	9	5
21–25	16	3
26–30	18	1
Total	46	16

Answer 3.87

	Disease	
Test	positive	negative
positive	12	12
negative	4	34

Table 4. True positives, true negatives, false positives and false negatives of the CMMS test.

$$sensitivity = \frac{12}{16} = 0.75$$

Answer 3.88

$$specificity = \frac{34}{46} = 0.74$$

Additional Question

$$False\ positive\ rate = \frac{12}{46} = 0.26$$

$$False\ negative\ rate = \frac{4}{16} = 0.25$$

$$\Pr(D) = \frac{46}{46+16} = 0.742$$

$$PV^+ = \Pr(D|T^+) = \frac{\Pr(T^+|D)\Pr(D)}{Pr(T^+|D)\Pr(D) + \Pr(T^+|\bar{D})\Pr(\bar{D})}$$
$$= \frac{0.75 \times 0.742}{0.75 \times 0.742 + 0.26 \times (1 - 0.742)} = 0.892$$

$$PV^- = \Pr(\bar{D}|T^-) = \frac{\Pr(T^-|\bar{D})\Pr(\bar{D})}{\Pr(T^-|\bar{D})\Pr(\bar{D}) + \Pr(T^-|D)\Pr(D)}$$
$$= \frac{0.74 \times (1 - 0.742)}{0.74 \times (1 - 0.742) + 0.25 \times 0.742} = 0.507$$

A study [12] of incidence rates of blindness among insulin-dependent diabetics reported that the annual incidence rate of blindness per year was 0.67% among 30- to 39-year-old male insulin-dependent diabetics (IDDM) and 0.74% among 30- to 39-year-old female insulin-dependent diabetics.

4.60 If a group of 200 IDDM 30- to 39-year-old women is followed, what is the probability that at least 2 will go blind over a 1-year period?

Answer 4.60

We have the binomial distribution with n=200, p=0.0074

$$\Pr(X \geq 2) = 1 - \Pr(X \leq 1) = 1 - [\Pr(X = 0) + \Pr(X = 1)]$$
$$= 1 - \left[\binom{200}{0} \times 0.0074^0 \times 0.9926^{200} + \binom{200}{1} \times 0.0074^1 \times 0.9926^{199}\right]$$
$$= 1 - (0.2264 + 0.3376) = 0.436$$

Obstetrics

The standard screening test for Down's syndrome is based on a combination of maternal age and the level of serum alpha-fetoprotein. Using this test 80% of Down's syndrome cases can be identified, while 5% of normals are detected as positive.

Suppose that 1 out of 500 infants are born with Down's syndrome.

10.124 What precentage of infants who test positive using the test will actually have Down's syndrome?

Answer 10.124

$$sensitivity = 0.8$$
$$false\ positive\ rate = 0.05$$

$$PV^+ = \Pr(D|T^+) = \frac{\Pr(T^+|D)\Pr(D)}{Pr(T^+|D)\Pr(D) + \Pr(T^+|\bar{D})\Pr(\bar{D})} = \frac{0.8 \times \frac{1}{500}}{0.8 \times \frac{1}{500} + 0.05 \times (1 - \frac{1}{500})}$$
$$= 0.031 = 3.1\%$$