Introductory Biostatistics for Biologists

IGC, September 11th - September 15th, 2017

Exercises V / Exercises VI / Exercises VII

Statistical Inference

1. Two different methods of analysis were used to determine the concentration of paracetamol (% m/m) in tablets. Ten tablets of ten different lots were used in the experiment, whose aim was to determine if both methods gave different results. In order to reduce variability both methods were used in each tablet. The results are shown in the following table:

M1	84.63	84.38	84.08	84.41	83.82	83.55	83.92	83.69	84.06	84.03
M2	83.15	83.72	83.84	84.20	83.92	84.16	84.02	83.60	84.13	84.24

   Can we say that the results produced by the two methods are different?

2. The population proportion of individuals under 40 diagnosed with lung cancer that survive at least five years is unknown. In a random sample of 260 such patients, 30 survive at least five years.

   (a) Give a justification for why statistical inference based on the normal distribution is appropriate.

   (b) Give a 95% confidence interval for the population five-year survival rate for this type of patient.

   (c) Test the hypothesis that the population five-year survival rate is 10% versus the alternative that it is higher. State hypothesis, calculate the test statistic and report a p-value. Summarize your findings in the context of the problem.

3. The body temperature data set "Data_P4" contains the body temperature (Fahrenheit) and the gender of 130 volunteers, 65 men and

65 women. Answer the following question justifying the distributions considered to perform the tests.

(a) Is the mean body temperature of human adults really 98.6$^{\text{O}}$F (37$^{\text{O}}$C)? (Ignore differences due to gender)

(b) Is the mean body temperature of women higher than the mean body temperature of men?

(c) Construct a 90% confidence interval for the difference of means for both genders. Use this interval to test the hypothesis of the difference of mean values, $\mu_{women} - \mu_{men}$, being equal to one.

4. The following data reports readings from 12 home radon detectors exposed to 105 picocuries per liter of radon:

   (91.9, 97.8, 111.4, 122.3, 105.4, 95.0, 103.8, 99.6, 96.6, 119.3, 104.8, 101.7).

   We would like to test hypotheses about the median reading from home radon detectors:

   $$H_0\text{: median}=105 \qquad \textit{vs.} \qquad H_1\text{: median}\neq 105$$

   To do this, apply the Wilcoxon signed-rank statistic to the differences between the observations and 105. What do you conclude?

5. The weight of 5 grains was measured from two experimental varieties designated *premier* and *super*. Each grain was measured in mg to one decimal place. It is intended to determine whether there are significant differences between the two weights varieties.

premier	24.5	23.4	25.3	23.4	22.1
super	26.4	27.0	25.2	25.8	27.1

   Given that the number of observations is very small, use a nonparametric test to answer the question.

6. Nurses in an inner-city hospital were unknowingly observed on their use of latex gloves during procedures for which glove use is recommended. The nurses then attended a presentation on the importance of glove use. One month after the presentation, the same nurses were observed again. Here are the percentagens of procedures for which each nurse were gloves:

Nurse	Before	After	Nurse	Before	After
1	0.857	0.500	8	0.000	1.000
2	0.500	0.833	9	0.000	0.667
3	1.000	1.000	10	0.167	1.000
4	0.000	1.000	11	0.000	0.750
5	0.000	1.000	12	0.000	1.000
6	0.000	1.000	13	0.000	1.000
7	1.000	1.000	14	1.000	1.000

(a) Why is a one-sided alternative proper here? Why must matched pairs methods be used?

(b) Does the test indicate that the presentation was helpful? Use an adequate hypothesis testing verifying the assumptions.

(c) Find a 90% confidence interval for the difference in mean values after and before the presentation.

7. We have the following R commands and results:

```
> x1=c(10,13,16,19,17,15,20,23,15,16)
> x2=c(13,16,20,25,18,16,27,30,17,19)
> t.test(x1,x2,alt="less",conf.level=0.95,var.equal = TRUE))
 Two Sample t-test
data:  x1 and x2
t = -1.779, df = 18, p-value = 0.04606
alternative hypothesis:  true difference in means is less than
0
95 percent confidence interval:
-Inf -0.09349972
sample estimates:
mean of x mean of y
16.4 20.1
```

Write down the null and the alternative hypothesis, α, n_1 and n_2 corresponding to this output. What is the estimated standard error of the difference between the mean values? What R command would you use to find the critical value for the hypothesis used? (The value at the boundary of the rejection region is called the critical value)

8. During an experiment using laboratory animals the following data on renal corticol blood flow (ml/g/min) during control conditions and during the admistration of a certain anesthetic were recorded.

Animal Number	Control	Anesthetic	Animal Number	Control	Anesthetic
1	2.35	2.00	9	2.58	2.10
2	2.55	1.71	10	2.66	2.58
3	1.95	2.22	11	2.31	1.32
4	2.79	2.71	12	3.43	3.70
5	3.21	1.83	13	2.37	1.59
6	2.97	2.14	14	1.82	2.07
7	3.44	3.72	15	2.98	2.15
8	2.53	2.05			

Can one conclude on the basis of these data that the anesthetic retards renal corticol blood flow?

9. Using the following data, test the null hypothesis that male and female turtles have the same mean serum cholesterol concentrations.

Serum Cholesterol (mg/100 ml)	
Male	Female
220.1	223.4
218.6	221.5
229.6	230.2
228.8	224.3
222.0	223.8
224.1	230.8
226.5	

10. Oats are an important source of fibre and nutrients but are often avoided by people with coeliac disease. Srinivasan et al. (1996) gave 10 pacients with the disease an "oats challenge": 50 g of oats per day for 12 weeks. Results for enterocyte heigth (μm) before and after the challenge were:

Patient	1	2	3	4	5	6	7	8	9	10
Before	36.3	36.0	40.8	44.9	32.8	28.8	38.4	31.1	29.8	30.2
After	35.3	38.3	37.9	37.6	28.8	27.1	42.6	34.7	30.6	36.8

(a) Test the null hypotesis that the oats challenge had no effect on enterocyte heigths.

(b) Find a 90% confidence interval for the difference in mean values after and before the challenge.

(c) What conclusion do you draw?

11. Consider the following contingency table with the observed frequencies of each of four hair colors and each of two sexes:

Sex	Black	Brown	Blond	Red
Male	32	43	16	9
Female	55	65	64	16

Test the hypothesis of independence between hair color and sex.

12. Consider the following data for the abundance of a certain species of bird:

Sex	Spring	Summer	Fall	Winter
Males	163	135	71	43
Females	86	77	40	38

Test the hypothesis that the proportion of males and females is the same in all four seasons.