

1. Briefly answer the following:

a. *When a data set is not normally distributed, what should you try in order to appropriately make statistical tests on that data?*

transformations

b. *Why are paired samples analyzed differently than separate samples?*

The two individuals in a pair share many things in common with each other but differ from members of other pairs. Whatever variation these shared differences causes in the response variable is factored out in the difference between them. By looking at the differences, we potentially avoid much of the error variance in the data. Separate samples do not share these properties.

c. *What are the non-parametric alternatives to*

ANOVA

Kruskall-Wallis

2-sample t-test

Mann-Whitney U test

correlation

Spearman's rank correlation

2. Scorpionfly males were raised on either a high nutrient (NH) or low nutrient (NL) food, to test the prediction that bad conditions causes more developmental problems (as measured by asymmetric development). Their wings were then measured on both sides, and the absolute value of the difference between the left wing length and the right wing length was used as a measure of their asymmetry ( $X_i$ ). The results were as follows:

|          | sample size | $\sum X_i$ | $\sum X_i^2$ |
|----------|-------------|------------|--------------|
| NH males | 29          | 0.9744     | 0.0427       |
| NL males | 34          | 1.8598     | 0.1727       |

Do the appropriate test of the hypothesis, making the necessary assumptions.

$H_0$ : NH males and NL males have on average the same level of asymmetric development.

$H_A$ : NH males and NL males have on average different levels of asymmetric development.

We'll do a two-sample  $t$ -test.

|          | mean   | $s^2$       |
|----------|--------|-------------|
| NH males | 0.0336 | 0.00035572  |
| NL males | 0.0547 | 0.002150574 |

$$s_p^2 = \frac{v_1 s_1^2 + v_2 s_2^2}{v_1 + v_2} = \frac{28(0.00035572) + 33(0.002150574)}{28 + 33} = 0.0013267$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{0.0013267}{29} + \frac{0.0013267}{34}} = 0.009207$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{0.0336 - 0.0547}{0.009207} = -2.29$$

$t_{0.05(2),61} = 2.00$ , so we can reject the  $H_0$ . The low nutrition males have more developmental problems than the high nutrition males.

3. In Vancouver, the probability of rain during a winter day is 0.58, for a spring day 0.38, for a summer day 0.25, and in fall the probability of rain per day is 0.53. Each of these seasons lasts one quarter of the year.

a. What is the probability of rain per day in Vancouver?

$$(0.25)(0.58) + (0.25)(0.38) + (0.25)(0.25) + (0.25)(0.53) = 0.435.$$

b. If you are told that a particular day was raining, what is the probability that that day was a winter day?

$$\frac{(0.25)(0.58)}{0.435} = 0.333$$

4. A cross of two duck species (mallards and pintails) was made to investigate the behavioral patterns of hybrids. A behavioural index was calculated for each of 11 male F2 hybrids, and a plumage index (to measure the condition of their feathers) was also calculated for each of the birds. Are the behavioral and plumage indices independent? Make, and state, any necessary assumptions.

| Male | Plumage Index | Behavioral Index |
|------|---------------|------------------|
| R    | 7             | 3                |
| S    | 13            | 10               |
| D    | 14            | 11               |
| F    | 6             | 5                |
| W    | 14            | 15               |
| K    | 15            | 15               |
| U    | 4             | 7                |

|   |    |    |
|---|----|----|
| O | 8  | 10 |
| V | 7  | 4  |
| J | 9  | 9  |
| L | 14 | 11 |

---

$$\begin{aligned}\sum P &= 111 \\ \sum P^2 &= 1277\end{aligned}$$

$$\begin{aligned}\sum B &= 100 \\ \sum B^2 &= 1072 \\ \sum PB &= 1141\end{aligned}$$

Correlation:

$H_0$ : Plumage index is not correlated with the behavioral index ( $\rho=0$ ).

$H_A$ : Slow wave sleep activity is correlated with the behavioral index ( $\rho \neq 0$ ).

Assumes a random sample, that all values of Y for a given X are normally distributed with equal variance for all values of X, and that all values of X for a given Y are normally distributed with equal variance for all values of Y.

$$\sum (X_i - \bar{X})^2 = \sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n} = 1277 - \frac{111^2}{11} = 156.9$$

$$\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n} = 1072 - \frac{100^2}{11} = 162.9$$

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} = 1141 - \frac{111(100)}{11} = 131.9$$

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = \frac{131.9}{\sqrt{(156.9)(162.9)}} = 0.825$$

$$SE[r] = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.825)^2}{11-2}} = 0.1884$$

$$t = \frac{r}{SE[r]} = \frac{0.825}{0.1884} = 4.38$$

We compare this to  $t_{0.05(2),9} = 2.26$ , to find that  $P < 0.05$ . In fact,  $P < 0.01$ . We can reject the null hypothesis and say that the plumage index and the behavioral index are positively correlated.

5. What is the best statistical technique for dealing with the following situations and questions? Be as specific as possible (i.e. “multivariate ANOVA” is a better answer than “ANOVA,” if it is a more appropriate technique.)

A. Are the nucleotides used in 2 regions of DNA composed of the same proportions of A, T, C and G?

contingency analysis

B. Do grades on homework assignments predict final exam grades?

regression

D. In humans, is hair length different on average between the sexes? (Assume normality and that the variance of hair length is equal.)

two-sample *t*-test

E. Re-answer D., if the distributions were not normal.

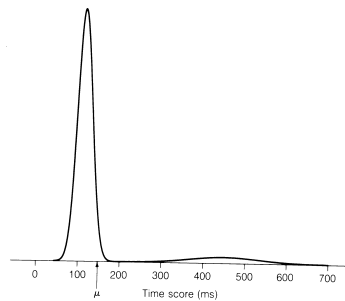
Mann-Whitney U test

F. Data on 412 married couples recorded the amount of weight gain in 10 years after marriage. Is there a difference in weight gain after marriage in men relative to women?

paired *t*-test

G. This graph presents the distribution of time required to push a button after a stimulus. Comparing 7 different groups, are all groups the same in their average button-pushing times?

FIGURE 5.14  
Distribution of Time  
Scores in a Button-  
Pushing Task



Kruskal-Wallis test

H. Is the number of chirps per hour of calling birds normally distributed?

Shapiro-Wilk test

I. A study compared the effects of 5 soil types and 6 different fertilizers on corn growth rate. Is the effect of soil type independent of the effect of fertilizer type?

two-factor ANOVA

J. Do men and women have the same variance of final grade averages (if final grades are normally distributed within sexes)?

F test

6. A biological statistics class was given a test. Immediately after the test was completed, each student was asked if she or he felt better or worse than before they took the test. Ten reported feeling better, and two felt worse. Using this sample of twelve students, does taking stats exams change the average mood of students?

$H_0$ : Stats tests are as likely to improve mood as to decrease mood.

$H_A$ : Stats tests change students' mood on average.

$$P = 2(\Pr[10] + \Pr[11] + \Pr[12])$$

$$\begin{aligned} P &= 2 \left[ \binom{12}{10} (0.5)^{10} (0.5)^2 + \binom{12}{11} (0.5)^{11} (0.5)^1 + \binom{12}{12} (0.5)^{12} (0.5)^0 \right] \\ &= 2[0.0161133 + 0.00292969 + 0.0002441] \\ &= 0.0386 \end{aligned}$$

Thus we can reject the null hypothesis. Student's moods are more likely to improve by stats tests than to get worse.

7. Monoamine oxidase (MAO) is an enzyme that is thought to play a role in the regulation of behavior. To see whether different categories of schizophrenic patients have different levels of MAO activity, researchers collected blood specimens from 42 patients and

measured the MAO activity in the platelets. The means and standard deviations of each group are summarized below. (Values are expressed as nmol benzylaldehyde product/ $10^8$  platelets/hour.) Construct the ANOVA table and test whether these behavioral categories have any differences in MAO activity. Assume that MAO activity is normally distributed and that each of these groups has equal variance in MAO activity.

| Diagnosis                               | Mean MAO Activity | SD MAO Activity | No. of patients |
|---|-------------------|-----------------|-----------------|
| Chronic undifferentiated schizophrenic  | 9.81              | 3.62            | 18              |
| Undifferentiated with paranoid features | 6.28              | 2.88            | 16              |
| Paranoid schizophrenic                  | 5.97              | 3.19            | 8               |

$H_0$ : The three groups all have the same mean MAO activity.

$H_A$ : At least one of the three groups has a different MAO activity on average.

| Source of variation | Sum of Squares | df | Mean Squares | F ratio | P     |
|---------------------|----------------|----|--------------|---------|-------|
| MAO Group           | 136.29         | 2  | 68.15        | 6.35    | 0.004 |
| Error               | 418.42         | 39 | 10.7287      |         |       |
| Total               | 554.71         | 41 |              |         |       |

The groups vary in their MAO mean activity.

8. List the assumptions of the following tests

a. *Paired t-test*

random sample of pairs, the differences are normally distributed

b. *Mann-Whitney U test*

random samples

c. *Linear regression*

random sample of Y values for each X. Y is normally distributed with equal variance for each value of X. The relationship follows a line.

d. *ANOVA*

All samples are random samples from populations that are normally distributed with equal variance.