Ma 322: Biostatistics
Homework Assignment 8

Prof. Wickerhauser

Read Chapter 14, "Hypothesis Testing," pages 240–262 of our text.

1. Following are 14 samples from a normal population with unknown mean and unknown standard deviation:

   ```
   2.68 5.11 3.66 0.87 4.34 4.12 3.79 2.05 2.59 2.56 4.10 1.99 3.46 1.97
   ```

   (a) Estimate the mean μ, the standard deviation σ, and the variance σ^2 from this sample.

   (b) Test the hypothesis $H_0 : \mu = 3.0$, using the significance level $\alpha = 0.05$.

   (c) Test the hypothesis $H_0 : \mu \leq 2.5$, using the significance level $\alpha = 0.05$.

2. Using the sample standard deviation from Exercise 1 and a significance level of $\alpha = 0.05$, determine:

   (a) The power $1 - \beta$ of the t-test to reject the two-sided null hypothesis on the mean in Exercise 1b when there is a true difference $\delta = 0.5$.

   (b) The power $1 - \beta$ of the t-test to reject the one-sided null hypothesis on the mean in Exercise 1c when there is a true difference $\delta = 0.5$.

   (c) The number of samples needed to get a power $1 - \beta = 99\%$ in the t-test of the two-sided null hypothesis on the mean in Exercise 1b when there is a true difference $\delta = 0.5$.

   (d) The number of samples needed to get a power $1 - \beta = 99\%$ in the t-test of the one-sided null hypothesis on the mean in Exercise 1c when there is a true difference $\delta = 0.5$.

3. (a) Using the following data, and assuming that both populations are normal with equal variance, test the null hypothesis that male and female turtles have the same mean serum cholesterol concentrations.

Serum cholesterol (mg/100 ml) of turtles.

Male	248,329,223,313,271,324,255,255,423,332,311,264
Female	341,311,362,371,419,366,246,273,312,331

(b) The following data were found in Table 1 of C. M. Holcomb, C. G. Jackson, Jr., and M. M. Jackson, "Serum Cholesterol Values in Three Species of Turtles," J. Wildlife Diseases 8(1972), pp.181–182. `<www.jwildlifedis.org/cgi/reprint/8/2/181.pdf>`

Serum cholesterol (mg/100 ml) in turtles.

Species	n	Mean	S.E.	Range	Coef. of Var.
C. scripta	8	290.0	±42.3	174–512	41.2%
T. carolina	31	339.7	±15.6	178–511	25.6%

Assuming that both populations are normal with equal variance, test the alternative hypothesis that *T. carolina* has higher mean serum cholesterol concentrations than *C. scripta*.

4. For a fair coin, expect a binomial distribution with "heads" probability $p = 1/2$. A certain guilder coin is tossed 2000 times and comes up heads just 962 times.

   (a) Rosencrantz does not believe that this guilder is a fair coin. Use the experimental data and a significance threshold of $\alpha = 0.05$ to test Rosencrantz's one-sided hypothesis H_A: heads are less likely than tails in a toss of that coin.

   (b) Guildenstern does not share Rosencrantz's suspicions about the coin. Use the experimental data and a significance threshold of $\alpha = 0.05$ to test Guildenstern's two-sided hypothesis H_0: heads and tails are equally likely in a toss of that coin.

5. (a) Using the data for Problem 3, part a, test the null hypothesis that male and female turtles have the same serum cholesterol variance.

   (b) Using the data for Problem 3, part b, test the alternative hypothesis that *C. scripta* has a higher serum cholesterol variance than *T. carolina*.

6. (a) Test the hypothesis that nucleotides a,c,g,t are equally likely in the GenBank sequence `NM_005369`, using the χ^2 goodness-of-fit method. Use significance level $\alpha = 0.01$.

   (b) Test the hypothesis that nucleotides a,c,g,t are equally likely in the GenBank sequence `NM_005367`, using the χ^2 goodness-of-fit method. Use significance level $\alpha = 0.01$.

7. (a) How many 2×2 contingency tables are there with row sums $(2, 5)$ and column sums $(3, 4)$? (Hint: Write down all the solutions.)

(b) Assuming that the rows and columns are independent, compute the exact hypergeometric probability of each 2×2 contingency table in part a.

8. The following data are frequencies of bats found with and without rabies in two different geographic areas:

Area	With rabies	Without rabies
E	11	112
W	18	139

(a) Using the Yates-corrected χ^2 test at the $\alpha = 0.05$ significance level, test H_0: the incidence of rabies is the same in both areas.

(b) Use the Fisher exact test at the 0.05 level to test if the E population bats are less likely to have rabies than those in the W population.

9. A follow-on study was performed on the same bats data, similar to that of Problem 8 but with the additional tabulation of gender:

	With rabies		Without rabies	
Area	Male	Female	Male	Female
E	6	5	49	63
W	14	4	84	55

(a) Test for mutual independence at the $\alpha = 0.05$ significance level.

(b) Test for partial independence at the $\alpha = 0.05$ significance level.