Statistics and Bioinformatics
Problem set 10
Due in class, Tuesday, December 21

1) Using qchisq() and assuming 10 degrees of freedom, find the value of the chi-square distribution for which
   a) 5% of the distribution is higher

      qchisq(0.95, 10) = 18.31

   b) 5% of the distribution is lower

      qchisq(0.05, 10) = 3.94

   c) 2.5% of the distribution is higher

      qchisq(0.975, 10) = 20.5

   d) 2.5% of the distribution is lower

      qchisq(0.025, 10) = 3.25

   e) 1% of the distribution is higher

      qchisq(0.99, 10) = 23.2

   f) 1% of the distribution is lower qchisq(0.01, 10) = 2.56

2) Using either qf() or the Vasilj textbook Tablica C1 (p. 300) find the value of the F distribution for which
   a) 5% of the distribution is higher, numerator and denominator degrees of freedom are 2 and 5

      qf(0.95, 2, 5) = 5.79

b) 1% of the distribution is higher, numerator and denominator degrees of freedom are 2 and 5

qf(0.99, 2, 5) = 13.27

c) 1% of the distribution is higher, numerator and denominator degrees of freedom are 50 and 100

qf(0.99, 50, 100) = 1.74

d) 1% of the distribution is higher, numerator and denominator degrees of freedom are 90 and 200

qf(0.99, 90, 200) = 1.50

e) 1% of the distribution is higher, numerator and denominator degrees of freedom are infinite and infinite

1

3) Given the values of F below and associated degrees of freedom, calculate the probability of the right tail (i.e. the probability that the F value is higher than that shown). Hint: use pf().

a) F = 2.3, numerator and denominator degrees of freedom are 3 and 7.

1 – pf(2.3, 3, 7) = 0.16

b) F = 1.2, numerator and denominator degrees of freedom are 3 and 745.

1 – pf(1.2, 3, 745) = 0.31

c) F = 3.12, numerator and denominator degrees of freedom are 20 and 221.

1 – pf(3.12, 20, 221) = 1.96×10^{-5}

d) F = 1.0, numerator and denominator degrees of freedom are 5 and 20.

1 – pf(1.0, 5, 20) = 0.44

e) F = 0.4, numerator and denominator degrees of freedom are 4 and 78.

1 – pf(0.4, 4, 78) = 0.81

4) A sample of plant mass is obtained, and the values measured are 5, 7, 4, 3, 3, 5, 4, 5, 7, 8.

a) Calculate the sample variance.

$s^2 = 2.99$

b) Calculate the 95% confidence interval for the population variance.

$high = \nu s^2 / \chi^2_{0.025} = (9)(2.99)/qchisq(0.025, 9) = 9.97$

$low = \nu s^2 / \chi^2_{0.975} = (9)(2.99)/qchisq(0.975, 9) = 1.41$

c) What does this calculation assume?

Assumes plant mass is normally distributed, and that these observations were a random sample.

5) The sample standard deviation (s) in photosynthesis rate was measured in n plants. Write the equation for the standard deviation of the sample mean (or SEM) in terms of s and n.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

6) Three random samples of n = 100 each were obtained, and a variable (x) was measured. The sample means of x were 3, 6, and 4.

a) What is the mean of the sample means?

$\bar{x} = 4.33$

b) What is the observed variance of the sample means?

$s^2_{\bar{x}} = 2.33$

c) How many degrees of freedom is this variance based on?

2

d) If the variance of x is 2, what is the expected variance in the sample means, assuming n = 100?

$$s_{\bar{x}}^2 = \frac{s_x^2}{n} = \frac{2}{100} = 0.02$$

e) Calculate the ratio of the observed to expected variance in the sample means.

$$\frac{2.33}{0.02} = 116.5$$

f) What probability distribution does this ratio follow if the null hypothesis is true?

F

g) What are the numerator and denominator degrees of freedom for this ratio?

numerator: 2

denominator: $3(n-1) = 297$

h) Perform an F-test of the hypothesis that the population means are equal (compare your observed F with the critical value of F). Do you reject the null hypothesis?

$F_{crit} = qf(0.95, 2, 297) = 3.03$

$F >> F_{crit}$ ∴ reject H_0, means are not all equal.

7) Honeybees perform a dance to communicate to the hive travel directions to a source of nectar. During the dance, they vocalize in soundbursts. The number of soundbursts per dance cycle is correlated with the distance to the nectar source. There is some speculation that different European subspecies of honeybees have different dialects. Below are data of the average number of soundbursts per dance cycle for three European dialects. In each case, the distance to the nectar source is 250 m. Is there any evidence in these data for the existence of different dialects in the different subspecies? (Hint: test the null hypothesis that the mean soundbursts is the same for each subspecies in an ANOVA: calculate the F and compare it to the critical F or calculate its probability, and state whether the null hypothesis was rejected).

Apis mellifera ligustic	*A. mellifera carnica*	*A. mellifera cacasica*
16.5	14.5	15.2
15	16.2	16.8
16.2	14.1	17.9
16.4	14	16
16.9	14.3	16.3
17.9	13.7	14.9
14.2	14.5	15.6
17.2	14.1	15.9
16	13.1	16.7
15.7	14.9	16.2
	13.5	16.1
	14.2	15.9
	14.4	17.5
	13.1	

	df	SS	MS	F	p
groups	2	35.9	18.0	22.9	4.97×10^{-7}
within	34	26.6	0.78		

H_0 rejected