Statistics and Bioinformatics
Problem Set 7
Due in class, Tuesday, November 30

1. The standard deviation of the t-distribution is __________ (smaller, greater) than the standard deviation of the Z-distribution.

2. By convention, the t-distribution is nearly equal to the Z-distribution if the sample size is greater than about ___30___.

3. As the sample size increases, the standard deviation of the t-distribution becomes __________ (greater/smaller).

4. In the Z-distribution the tail containing 0.025 of the distribution (both tails add up to 0.05) corresponds to a Z value of about __________. qnorm(0.0975) = 1.96

5. In the t-distribution, the tail containing 0.025 of the distribution (both tails add up to 0.05) corresponds to a t value __________ (greater than, less than) 1.96.

6. If the sample size is 10, the t-value corresponding to a tail probability of 0.025 (both tails add up to 0.05) is ___2.26___. (Hint: use qt() in R, or look up this number in tablica B, p. 298 of Vasilj.)

   qt(0.0975, 9) = 2.26

7. If the sample size is 15, the t-value corresponding to a tail probability of 0.05 (both tails add up to 0.10) is ___1.76___. (Hint: use qt() in R, or look up this number in tablica B, p. 298 of Vasilj.)

   qt(0.95, 14) = 1.76

8. In a hypothesis test, you have two competing hypotheses about a population __________ (parameter, statistic).

9. In a hypothesis test, you test only the __________ (alternative, null) hypothesis.

10. A very important requirement for formulating the hypotheses for a hypothesis test is that no information from the __________ (population, sample) is used in the hypotheses statements.

11. The null hypothesis *always* contains the __________ (equality, inequality).

12. If the alternative hypothesis contains a > symbol, it is a __________ (right-tailed, left-tailed, two-tailed) test.

13. If the alternative hypothesis contains a < symbol, it is a _________ (right-tailed, left-tailed, two-tailed) test.

14. If the alternative hypothesis contains a $\neq$ symbol, it is a _________ (right-tailed, left-tailed, two-tailed) test.

15. The symbol $t_{alpha/2}$ is the value of the t distribution for which the tail probability is __________ (alpha, alpha/2)

16. The the null hypothesis is rejected if the absolute value of the observed t value (| t |) is greater than $t_{alpha/2}$ in the __________ (two-tailed, right-tailed, left-tailed) test.

17. The the null hypothesis is rejected if the observed t value is ________ (greater than, less than) ________ ($t_{alpha/2}$, t_{alpha}) in the right-tailed test.

18. For each of the following questions: 1) define the pertinent true means; 2) write down equations or inequalities that correspond to the null hypothesis and the alternative hypothesis, and label the two as null or alternative; 3) state whether the test is a left-tail, right-tail, or two-tailed test; 4) write the equation for t or Z that will be used to test the null hypothesis; and 5) write the criterion for rejection of the null hypothesis.

a) Is the sex ratio of males to females 1:1 in some population?

p = true proportion ♀

$H_0: p = 0.5$

$H_A: p \neq 0.5$

2-tailed

$$Z = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5^2}{n}}}$$

$|Z| > 1.96$

b) Does fertilizer increase the yield of some crop variety?

μ = true difference in yield

$H_A: \mu > 0$

$H_0: \mu \leq 0$

right-tailed

$$t = \frac{\bar{X} - 0}{\frac{S}{\sqrt{n}}}$$

$t > t_{crit}$

$t >$ qt(0.95, df)   ← n-1

c) Does a particular pesticide reduce the disease level of a particular crop?

μ = diff in disease level

$H_A: \mu < 0$

$H_0: \mu \geq 0$

left-tailed

$$t = \frac{\bar{X} - 0}{S/\sqrt{n}}$$

$t < t_{crit}$

$t <$ -qt(0.95, df)   ← n-1

In these 2-sample t-tests below, you can assume either equal variances in the two populations, or unequal variances. I've assumed unequal variances below. In that case the degrees of freedom can be the smaller of n1 - 1 and n2 - 1. Or it can be the Satterthwaite number that R uses in the t.test() function. If you assume equal variances, you calculate the pooled standard deviation, and the degrees of freedom is n1 + n2 - 2.

d) Does crop variety A have a higher yield than crop variety B?

μ_A = true yield A, μ_B = true yield B

$H_A: \mu_A - \mu_B > 0$ right-tailed

$H_0: \mu_A - \mu_B \le 0$

$$t = \frac{\bar{x}_A - \bar{x}_B - 0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

$t > t_{crit}$

$t > qt(0.95, df)$

e) Does fertilizer A have the same effect on yield as fertilizer B?

μ_A = true effect of A, μ_B = true effect of B

$H_0: \mu_A - \mu_B = 0$ 2-tailed test

$H_A: \mu_A - \mu_B \neq 0$

$$t = \frac{\bar{x}_A - \bar{x}_B - 0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

$|t| > t_{crit}$

$|t| > qt(0.975, df)$

f) Does variety A of a stock animal have the same rate of food consumption as variety B?

μ_A = true rate of consumption of A

μ_B = true rate of consumption of B

$H_0: \mu_A = \mu_B$ or $\mu_A - \mu_B = 0$ 2-tailed

$H_A: \mu_A \neq \mu_B$ or $\mu_A - \mu_B \neq 0$

$$t = \frac{\bar{y}_A - \bar{x}_B - 0}{\sqrt{s_A^2/n_A + s_B^2/n_B}}$$

$|t| > t_{crit}$

$|t| > qt(0.975, df)$

g) Does soil type A produce bigger tomatoes than soil type B?

μ_A = true size tomatoes in A

μ_B = " " " " B

right-tailed

$H_A: \mu_A - \mu_B > 0$

$H_0: \mu_A - \mu_B \le 0$

$$t = \frac{\bar{x}_A - \bar{x}_B - 0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

$t > t_{crit}$

$t > qt(0.95, df)$

h) Is the level of arsenic in the water supply higher than some value x?

μ = true level of arsenic

right-tailed

$H_A: \mu - x > 0$

$H_0: \mu - x \leq 0$

here x is a constant

$$t = \frac{\bar{x} - x}{s/\sqrt{n}}$$

$t > t_{crit}$

$t > qt(0.95, \underset{n-1}{df})$

i) Is the mutation rate in region A of the genome the same as in region B of the genome of a particular species?

μ_A = true mutation rate in A

μ_B = " " " " B

$H_0: \mu_A - \mu_B = 0$

2-tailed

$H_A: \mu_A - \mu_B \neq 0$

$$t = \frac{\bar{x}_A - \bar{x}_B - 0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

$|t| > t_{crit}$

$|t| > qt(0.975, df)$

j) Is a genetic trait shown in 75% of the progeny of a particular cross?

p = true proportion progeny with trait

2-tailed

$H_0: p = 0.75$

$H_A: p \neq 0.75$

$$z = \frac{\hat{p} - 0.75}{\sqrt{\frac{(0.75)(0.25)}{n}}}$$

$|z| > z_{crit}$

$|z| > qnorm(0.975)$

$|z| > 1.96$

k) Does infection by a particular virus increase the mortality of a plant variety?

μ = true mean difference in mortality

right-tailed test

$H_A: \mu > 0$

$H_0: \mu \leq 0$

$$t = \frac{\bar{x} - 0}{\frac{s}{\sqrt{n}}}$$

$t > t_{crit}$

$t > qt(0.95, \underset{n-1}{df})$

l) Does a particular lowfat diet actually work (i.e., does it cause people to lose weight)?

μ = true mean weight change

left-tailed

$H_A: \mu < 0$

$H_0: \mu \geq 0$

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}}$$

$t < t_{crit}$

$t < -qt(0.95, \underset{n-1}{df})$

Note: the critical t is negative in a left tailed test.

m) Does a particular medication actually work (i.e., does it reduce the symptoms of a disease)?

μ = true mean difference in symptoms

left-tailed test

$H_A: \mu < 0$

$H_0: \mu \geq 0$

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}}$$

$t < t_{crit}$

$t < -qt(0.95, \underset{n-1}{df})$

19. If the alpha level of a test is 0.05, and the test is two-tailed, what is the size of each tail?

0.025

20. If the alpha level of a test is 0.05, and the test is right-tailed, what is the size of the right tail?

0.05

21. If the alpha level of a test is 0.05, the test is left-tailed, and the sample size is 50, what is the critical t-value for the test?

$t_{crit} = -qt(0.95, 49) = -1.68$

$= -qt(1 - \alpha/T, df)$

$df = n-1$
$= 50-1$

$\alpha = 0.05$

$T = 1$ (1-tailed)

22. If the alpha level of a test is 0.05, the test is two-tailed, and the sample size is 30, what is the critical t-value for the test?

$t_{crit} = qt(1-\alpha/T, df) = qt(1-0.05/2, 29)$

$= qt(1-0.025, 29)$

$= qt(0.975, 29) = 2.05$

$\left(\begin{array}{l} \alpha = 0.05 \\ T = 2 \text{ tails} \\ df = n-1 = 30-1 = 29 \end{array} \right)$

23. A sample of 100 volunteers agree to an exercise program and monitor their weight gain or loss after six months. The sample mean weight change per person was -0.5 kg, and the sample standard deviation in weight gain was 2 kg. The research question was: Does the exercise program cause the average person to lose weight? Write down the null hypothesis, the alternative hypothesis, the type of test (two-tailed, right tailed, left tailed), the criterion for rejection of the null hypothesis, the observed t-value for the test and its associated probability, and conclude whether the null hypothesis was rejected.

μ = true mean weight change

$H_A: \mu < 0$ left-tailed

$H_0: \mu \geq 0$

$t < t_{crit}$

$t < -qt(0.95, \underset{n-1}{df})$

$t_{crit} = qt(0.95, 99)$

$= -1.66$

$-2.5 < -1.66$

$\therefore$ reject H_0

$p = pt(-2.5, 99)$

$= 0.0070$

$$t = \frac{-0.5 - 0}{\frac{2}{\sqrt{100}}} = \frac{(-0.5)(10)}{2} = -5/2 = -2.5$$

24. For the question above, calculate the 95% confidence interval for weight gain.

$$95\% CI = -0.5 \pm (1.66)\left(\frac{2}{\sqrt{100}}\right)$$

$$= -0.5 \pm 0.332$$

$$= (-0.168, -0.832) \text{ kg}$$

25. Our first exam was taken by 14 women and 13 men. The mean female score was 72.7, and the mean male score was 70.3. The standard deviation in score was 7.7 for females and 8.3 for males.

a) Assume the true (population) standard deviation is the same in males and females. Calculate the pooled standard deviation, the t-value, and its tail probability. Was there a significant difference between males and females in exam scores? What kind of test is this (2-tailed, right-tailed, left-tailed)?

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(13)(7.7^2) + 12(8.3^2)}{14+13-2}} = 7.99$$

2-tailed test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{2.4}{7.99\sqrt{\frac{1}{14}+\frac{1}{13}}} = \frac{2.4}{3.08} = 0.779$$

$$t_{crit} = qt(0.975, 25) = 2.06$$

$t < t_{crit}$ ∴ not significant

b) Assume the true (population) standard deviations are different in males and females. Calculate the t-value and its tail probability. How many degrees of freedom did you use for this test? Was there a significant difference between males and females in exam scores? What kind of test is this (2-tailed, right-tailed, left-tailed)?

2-tailed

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2.4}{\sqrt{\frac{7.7^2}{14} + \frac{8.3^2}{13}}} = \frac{2.4}{3.09} = 0.777$$

not significant

26. Two varieties of pea plants, A and B, are exposed to a virus. The number of plants exposed in each variety was 500. After two weeks, the number of A plants found to be infected was 140. The number of B plants found to be infected was 105. Is there a true difference in the mean resistance to the virus within the two populations of pea varieties? Show your observed test statistic and its tail probability. What kind of test is this (2-tailed, right-tailed, left-tailed)?

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\bar{p}\bar{q}}{n_A} + \frac{\bar{p}\bar{q}}{n_B}}}$$

$$\bar{p} = \frac{x_A + x_B}{n_A + n_B} = \frac{140+105}{500+500}$$

$$\bar{p} = \frac{245}{1000} = 0.245$$

$$\bar{q} = 1 - \bar{p} = 0.755$$

$$\hat{p}_A = 140/500 = 0.28$$

$$\hat{p}_B = 105/500 = 0.21$$

$$z = \frac{0.28 - 0.21}{\sqrt{\frac{(0.245)(0.755)}{500} + \frac{(0.245)(0.755)}{500}}}$$

$$z = \frac{0.07}{0.027} = 2.59$$

$|z| > z_{crit} = 1.96$

$\therefore$ reject H_0.

Conclude there is a true diff.