STAT 6200 — Design of Experiments for Research Workers
Lab 4 – Due Tuesday, October 26

Example — Salt Intake Among Hypertensives:

- To investigate whether people suffering from hypertension have different average daily salt intake than the general population, the average daily salt intake was measured for a sample of 10 hypertensives. The data are as follows:

$$92.8, 54.8, 51.6, 61.7, 50.8, 84.5, 34.7, 62.2, 11.0, 39.1$$

- These data are not normally distributed. However, the natural logarithm of salt intake is approximately normally distributed. The sample mean of the natural log salt intake is $\bar{x} = 3.871$ with a sample sd of $s = .599$.

- It is known that the average salt intake in the general population (on the natural log scale) is 1.8 with sd 1.1.

  Do hypertensives have elevated salt intake relative to the general population?

1. First, suppose we are willing to assume that the population sd σ for hypertensives is the same as in the general population. That is, suppose $\sigma = 1.1$. Let μ denote the population mean salt intake for hypertensives.

   We wish to test

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu > \mu_0, \quad \text{where } \mu_0 = 1.8$$

   – We use a one-sided alternative here because we wish to determine whether the salt intake is "elevated" rather than "different" from the general population.

   – Let's test at significance level $\alpha = .01$ here, just to make things interesting

   – Here, since we assume $\sigma = 1.1$ is known, we can use a z test. That is, our test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

   and we reject H_0 at level α if $z > z_{1-\alpha}$.

   – Computing our test statistic, we get

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.871 - 1.8}{1.1/\sqrt{10}} = 5.95$$

The relevant critical value is

$$z_{1-\alpha} = z_{1-.01} = z_{.99}$$

To get this value in Minitab, select

Calc → Probability Distributions → Normal...

then select "Inverse cumulative probability", set the "Mean:" to 0, set the "Standard deviation:" to 1, select "Input constant:", and enter the constant .99. Hitting OK yields the answer:

$$z_{.99} = 2.32635$$

Since $|z| = |5.95| = 5.95$ exceeds this critical value (by a lot), we reject H_0 and conclude that hypertensives do have eleveated mean salt intake relative to the general population.

- This result was obtained on the natural log scale, so, strictly speaking, we've established that hypertensives have elevated mean log salt intake relative to the general population. But if their mean log salt intake is higher then their mean salt intake must also be higher, so the result applies on the original scale as well.

  – The p-value for this test is

  $$p = P(Z > 5.95)$$

  which can be obtained in Minitab by selecting

  Calc → Probability Distributions → Normal...

  The select "Cumulative probability", set the mean and SD to 0 and 1, respectively, select "Input constant:", enter 5.95, and hit OK.

  These steps give

  $$P(Z \leq 5.95) = 1.00000$$

  Of course, this is a rounded figure. The true value must be somewhat less than 1, but equals 1.00000 when rounded to 5 decimal places.

  The p-value for this test is therefore

  $$p = P(Z > 5.95) = 1 - P(Z \leq 5.95) = 1 - 1.00000 = 0.00000$$

  which would normally be reported as $p < .0001$ or something similar.

- This whole procedure can be done in a more automated way in Minitab. Here are the steps: select

  Stat → Basic Statistics → 1-Sample Z...

Then under "Samples in columns:" select column C2, ln(SaltIntake). Then set "Standard deviation:" to 1.1, and set "Test mean:" to 1.8. Then go to "Options...", set "Confidence level:" to .99, set "Alternative:" to "greater than", and hit OK twice. The results of the test are displayed in the output window and agree with those obtained above.

- Note also that a 99% one-sided (lower) confidence bound is displayed. There is a direct correspondence between this bound and the test.
  - Specifically, all values of μ_0 that fall below this $100(1-alpha) = 99\%$ lower bound would lead to rejection of $H_0 : \mu = \mu_0$ versus $H_A : \mu > \mu_0$ at $\alpha = .01$
  - That is, the boundary of the rejection region for an α-level test of $H_0 : \mu = \mu_0$ versus $H_A : \mu > \mu_0$ forms a $100(1-\alpha)\%$ lower bound for μ.
  - Similarly, the boundary of the rejection region for an α-level test of $H_0 : \mu = \mu_0$ versus $H_A : \mu < \mu_0$ forms a $100(1-\alpha)\%$ upper bound for μ.
- In the two-sided case, there is also such a correspondence between confidence intervals and tests.
  - The acceptance region of an α-level test of $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$ forms a $100(1-\alpha)\%$ confidence interval for μ.

2. Now suppose that we are not willing to assume that the population sd among hypertensives is the same as that among the general population. Therefore, we must estimate σ rather than assume we know it. Repeat the test from part 1 under this scenario.
   - The null and alternative hypotheses remain the same:
   $$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu > \mu_0, \quad \text{where } \mu_0 = 1.8$$
   - The level of our test also remains the same: $\alpha = .01$.
   - However, now that we no longer know σ, we must estimate it. Therefore, our test statistic changes from z to $t = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$. We reject H_0 if $|t| > t_{1-\alpha}(n-1)$, or equivalently if $p = P(t(n-1) > t)$ is greater than α.
   - Our test statistic is computed as
   $$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.871 - 1.8}{.599/\sqrt{10}} = 10.93$$

- The relevant critical value is

$$t_{1-\alpha}(n-1) = t_{1-.01}(9) = 2.82144$$

which can be obtained in Minitab by selecting

Calc $\rightarrow$ Probability Distributions $\rightarrow$ t...

then select "Inverse cumulative probability", set the "Degrees of freedom:" to 9, select "Input constant:", and enter the constant .99. Hitting OK yields the answer.

- Alternatively, we can compute the p-value as

$$p = P(t(n-1) > t) = P(t(9) > 10.93) = 1-P(t(9) \leq 10.93) = 1-1.00000 = 0.00000$$

where $P(t(9) \leq 10.93)$ is obtained by selecting

Calc $\rightarrow$ Probability Distributions $\rightarrow$ t...

then select "Cumulative probability", set the "Degrees of freedom:" to 9, select "Input constant:", and enter the constant 10.93. Hitting OK yields the answer.

- Either way, we reject H_0 and conclude that hypertensives have elevated salt intakes.

- Again, this procedure can be automated in Minitab. Select

Stat $\rightarrow$ Basic Statistics $\rightarrow$ 1-Sample t...

Then under "Samples in columns:" select column C2, ln(SaltIntake). Then set "Test mean:" to 1.8, go to "Options...", set "Confidence level:" to .99, set "Alternative:" to "greater than", and hit OK twice. The results of the test are displayed in the output window and agree with those obtained above.

Exercise:

The Bayley Scales of Infant Development yield score on two indices — the Psychomotor Development Index (PDI) and the Mental Development Index (MDI) — which can be used to assess a child's level of functioning in each of these areas at approximately one year of age. Among normal healthy infants, both indices have a mean value of 100. As part of a study assessing the development and neurologic status of children who have undergone reparative heart surgery during the first three months of life, the Bayley Scales were administered to a sample of one-year-old infants born with congenital heart disease. The data are contained in the Minitab file heart.mtw on the course web page. Download this Minitab worksheet, open it in Minitab and perform the following exercises.

3. Suppose that we wish to test whether infants with congenital heart disease have a reduced PDI relative to healthy infants.

   a. What are the appropropriate null and alternative hypotheses here?

Answer: $H_0 : \mu = \mu_0$ *versus* $H_A : \mu < \mu_0$, *where* $\mu_0 = 100$.

   b. Compute the test statistic and critical value for an $\alpha = .01$ level test. State the conclusion of your test.

Answer: *The sample mean and sample SD of PDI here are* $\bar{x} = 94.7832$ *and* $s = 15.8510$, *respectively, so our test statistic is*

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{94.7832 - 100}{15.8510/\sqrt{143}} = -3.94.$$

The appropriate critical value is

$$t_{1-\alpha}(n-1) = t_{.99}(142) = 2.353$$

Since $\bar{x}$ *is less than* μ_0 *(which provides eveidence against* H_0 *and in favor of* $H_A : \mu < 100$*), and since* $|t| = 3.94 > t_{.99}(142) = 2.353$, *we reject* H_0, *and conclude that* μ *the population mean PDI for infants with congenital heart disease, is less than 100, the population mean in the healthy population.*

   c. Your critical value from part (b) should have been based upon a t distribution. Compute what the critical value would have been from a z (standard normal) distribution. Why are these critical values so close?

Answer: $z_{.99} = 2.32635$, *which is pretty close to 2.353, the critical value based on the* t *distribution. These values are close because the* t *distribution becomes more and more similar to the* z *distribution as the sample size increases. for* $n - 1 = 143 - 1 = 142$ *degrees of freedom, the* $t(142)$ *and* $N(0,1)$ *distributions are almost identical. A consequence of this is that, although the appropriate test here is a* t *test because the population SD is unknown, a* z *test would give an almost identical answer.*

   d. Use Minitab to obtain a 99% upper bound on μ, the mean PDI for infants with congenital heart disease. Is $\mu_0 = 100$ above or below that upper bou7nd? How is the answer to this questions related to your result from part (b)?

Answer: *The pp% upper bound is*

$$\bar{x} + t_{1-\alpha}(n-1)s/\sqrt{n} = 94.7832 + 2.353(15.8510)/\sqrt{143} = 97.902$$

   e. What is the p-value associated with your test?

Answer: *The* p *value is*

$$p = P(t(n-1) > |t|) = P(t(142) > 3.94) = .000065$$