STA 100 Solutions HW 4

Dr. Erin K. Melcon

1. Let Y = mens height. Then, $\bar{Y} \sim N(\mu = 69.7, \sigma = 2.8/\sqrt{n})$.
   (a) $Pr\{\bar{Y} > 72\} = Pr\{Z > (72 - 69.7)/(2.8/\sqrt{10})\} = P\{Z > 2.6\} = 1 - P\{Z \leq 2.6\} = 1 - 0.9953 = 0.0047$
   (b) $Pr\{Y > 72\} = Pr\{Z > (72 - 69.7)/(2.8)\} = P\{Z > 0.82\} = 1 - P\{Z \leq 0.82\} = 1 - 0.7939 = 0.2061$
   (c) The 80^{th} percentile for a z-score is 0.84. The corresponding average height is: $\bar{Y}_{80} = Z(\sigma/\sqrt{n}) + \mu = 70.4438$.
   (d) $Pr\{70 < \bar{Y} < 71\} = Pr\{0.34 < Z < 1.47\} = Pr\{Z \leq 1.47\} - Pr\{Z \leq 0.34\} = 0.9292 - 0.6331 = 0.2961$

---

2. (a) $Pr\{210 < Y < 220\} = Pr\{\frac{210-215}{15} < Z < \frac{220-215}{15}\} = Pr\{-0.33 < Z < 0.33\} = Pr\{Z \leq 0.33\} - Pr\{Z \leq -0.33\}$ $= 0.6293 - 0.3707 = 0.2586$
   (b) $Pr\{210 < \bar{Y} < 220\} = Pr\{\frac{210-215}{15/\sqrt{25}} < Z < \frac{220-215}{15/\sqrt{25}}\} = Pr\{-1.67 < Z < 1.67\} = Pr\{Z \leq 1.67\} - Pr\{Z \leq -1.67\}$ $= 0.9525 - 0.0475 = 0.905$
   (c) The 75^{th} percentile for a z-score is 0.67. The corresponding average height is: $\bar{Y}^{75} = Z^{75}(\sigma/\sqrt{n}) + \mu = (0.67)(15/\sqrt{25}) + 215 = 217.01$.
   (d) $Pr\{\bar{Y} < 210 | \bar{Y} < 220\} = \frac{Pr\{(\bar{Y}<210) \text{ and } (\bar{Y}<220)\}}{Pr\{\bar{Y}<220\}} = \frac{Pr\{\bar{Y}<210\}}{Pr\{\bar{Y}<220\}} = \frac{P\{Z<-1.67\}}{Pr\{Z<1.67\}} = \frac{0.0475}{0.9525} = 0.0499$

---

3. Note, we round the d.f. $n - 1$ down to the next lowest, which is 40.
   (a) The 90% confidence interval is calculated by: $\bar{y} \pm t_{\alpha/2}(s/\sqrt{n})$ or $28.86 \pm (1.684)(4.24/\sqrt{50})$ or (27.8502, 29.8698). Thus, the lower bound is : 27.8502
   (b) From (a), the upper bound is: 29.8698
   (c) We are 90% confident that the true average length of womens menstrual cycles is between 27.8502 and 29.8698 days.
   (d) Since the value 29.5 is in the confidence interval, our confidence interval does support the hypothesis.
   (e) Yes. A random sample was taken and the sample size is over 30.

---

4. Note, we round the d.f. $n - 1$ down to the next lowest, which is 60.
   (a) The 95% confidence interval is calculated by: $\bar{y} \pm t_{\alpha/2}(s/\sqrt{n})$ or $4.36 \pm (2)(0.42/\sqrt{66})$ or (4.2566, 4.4634). Thus, the lower bound is : 4.2566
   (b) From (a), the upper bound is: 4.4634
   (c) We are 95% confident that the true average serum potassium concentration for healthy women is between 4.2566 and 4.4634 mEq/l.
   (d) Since both bounds for the confidence interval are about 2.3, yes it does support the claim.
   (e) It would widen, since the value of $t_{\alpha/2}$ would increase. Or, because we have to cover more possibilities, the interval would widen.

---

5. Recall we use the row infinity, or ∞ for calculating sample sizes.
   (a) $n = \frac{t^2_{\alpha/2}s^2}{e^2} = \frac{(1.645)^2(4.24)^2}{(0.5)^2} = 194.5913 = 195$
   So we would need at least 195 people.
   (b) $n = \frac{t^2_{\alpha/2}s^2}{e^2} = \frac{(1.645)^2(4.24)^2}{(0.1)^2} = 4864.7835 = 4865$
   So we would need at least 4865 people.
   (c) As the margin of error decreases, we need to be more and more accurate, so the sample size must increase (as larger sample sizes lead to less error).

(d) As the standard deviation increases, the sample size needed tends to increase, since we need more people for the same amount of error (also because s^2 is in the numerator).

---

6. (a) The 99% confidence interval is: $(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, or $(490 - 500) \pm (2.611)\sqrt{\frac{32^2}{210} + \frac{48^2}{180}}$ or $(-10) \pm (2.611)(4.2043)$ or (-20.9774, 0.9774).
Thus, the lower bound is : -20.9774

(b) From (a), the upper bound is: 0.9774

(c) We are 99% confident that there is no significant difference in the true average costs for the two treatments.

(d) No, since the interval covers 0 there is no statistical evidence to suggest that the true average costs are different.

(e) It would be wider, since the value of $t_{\alpha/2}$ would increase (or because we have to cover more possibilites, and are less willing to draw "from the tails").

---

7. (a) The 95% confidence interval is: $(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, or $(36.93 - 31.36) \pm (2.064)\sqrt{\frac{4.23^2}{15} + \frac{3.35^2}{25}}$ or $(5.57) \pm (2.064)(1.2813)$ or (2.9254, 8.2146).
Thus, the lower bound is : 2.9254

(b) From (a), the upper bound is: 8.2146

(c) We are 95% confident that Brand A infants have higher true average weight gain than brand B, by at between 2.9254 ounces and 8.2146 ounces.

(d) Yes, since both values of the confidence interval are positive, this would suggest that $\mu_1 > \mu_2$.

(e) It would narrow, since there would be less error involved (the standard error would decrease).

---

8. (a) The null hypothesis is: $H_0 : \mu_1 \leq \mu_2$.

(b) The alternative is $H_A : \mu_1 > \mu_2$.

(c) The test-statistic is $t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(90-88)-0}{\sqrt{\frac{5.2^2}{65} + \frac{6.3^2}{74}}} = \frac{2}{0.9759} = 2.0494$

(d) Since on the t-table at d.f. = 100, we know that $Pr\{t > 1.984\} = 0.025$, and $Pr\{t > 2.081\} = 0.02$, our p-value has the range : $0.02 <$ p-value < 0.025 or $0.02 <$ p-value < 0.025.

(e) The probability of seeing our sample data (or more extreme) if in reality the true average pulse rate for smokers was less than or equal to that of non-smokers is between 0.02 and 0.025.

(f) Since the p-value is larger than our specified α value, we fail to reject H_0 at 1% signficance.

(g) We cannot support the claim that smokers have a higher pulse rate on average than non-smokers.

---

9. (a) The null hypothesis is: $H_0 : \mu_1 = \mu_2$

(b) The alternative is $H_A : \mu_1 \neq \mu_2$

(c) The test-statistic is $t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(3.2-4)-0}{\sqrt{\frac{0.9^2}{60} + \frac{0.7^2}{27}}} = \frac{-0.8}{0.1779} = -4.4969$

(d) Since on the t-table at d.f. = 60, we know that $Pr\{t > 3.46\} = 0.0005$, our p-value has the range :
p-value $< 0.0005 * 2$ or p-value < 0.001.

(e) The probability of seeing our sample data (or more extreme) if in reality the true average tail lengths for redbacked and ledbacked salamanders were equal is less than 0.001.

(f) Since the p-value so small, we would reject the null hypothesis at any reasonable value of α.

(g) We conclude at 1% significance that there is a difference in average tail lengths between the redbacked and leadbacked salamanders.
