Class # 25: Comparing two Population Means

Reading: Section 9.1 of Devore

At this point, it is easy for us to figure out how to compare the means of two populations. We begin by assuming that we sample \( n \) values from population 1 and \( m \) values from population 2. We will also assume that the sample from population 1 is normally distributed, i.e.,

\[ X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N \left( \mu_1, \sigma_1^2 \right), \]

and that the sample from population 2 is normally distributed, i.e.,

\[ Y_1, Y_2, \ldots, Y_m \overset{i.i.d.}{\sim} N \left( \mu_2, \sigma_2^2 \right). \]

If this is the case, it follows that \( \bar{X} \sim N \left( \mu_1, \sigma_1^2 / n \right) \) and \( \bar{Y} \sim N \left( \mu_2, \sigma_2^2 / m \right) \), and assuming independence, we get that

\[ \bar{X} - \bar{Y} \sim N \left( \mu_1 - \mu_2, \sigma_1^2 / n + \sigma_2^2 / m \right). \]

It is then easy to derive a \( 100(1 - \alpha)\% \) confidence interval for \( \mu_1 - \mu_2 \). Using methods identical to the ones given before, it is easy to show that this confidence interval is calculated as

\[ \bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}. \]

And if \( n \) and \( m \) are large enough, and if populations 1 and 2 are not necessarily normally distributed, then the distribution of \( (\bar{X} - \bar{Y}) / \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \) can be approximated by a standard normal, and an approximate \( 100(1 - \alpha)\% \) confidence interval for \( \mu_1 - \mu_2 \) can be calculated as

\[ \bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}. \]

And to test the hypothesis \( H_0 : \mu_1 - \mu_2 = \Delta_0 \), one proceeds just as outlined in the notes for Class 22. If both populations 1 and 2 are normally distributed and \( \sigma_1^2 \) and \( \sigma_2^2 \) are known, then the \( z \)-score is calculated as

\[ z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}, \]

and if both populations 1 and 2 are not normally distributed, but \( n \) and \( m \) are large enough, then the \( z \)-score is calculated as

\[ z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}. \]

Homework: These problems are due at the beginning of class on 4/26/10.

2. Devore, page 310, #39.
3. Devore, page 323, #74.
4. Devore, page 323, #78.
Homework Solutions:

1. $H_o : p = \frac{2}{3}$ vs. $H_a : p \neq \frac{2}{3}$. We calculate $z = (\hat{p} - p_o)/\sqrt{\frac{p_o(1-p_o)}{n}}$ where $\hat{p} = \frac{80}{124} = .645$. Therefore, $z = \frac{(\frac{2}{3} - 2/3)}{\sqrt{\frac{2/3 \times 1/3}{124}}} = -.511$. The $p$-value is calculated as $P(Z < -.511) + P(Z > .511)$ where $Z \sim N(0,1)$. Note that this $p$-value can be calculated as $2P(Z < -.511) = 2\text{pnorm}(-.511) = .61$. At the $\alpha = .05$ level, we would fail to reject $H_o$.

2. $H_o : p = .25$ vs. $H_a : p < .25$. In this case, $\hat{p} = \frac{177}{1050} = .168$. We calculate $z = (\hat{p} - p_o)/\sqrt{\frac{p_o(1-p_o)}{n}} = (\frac{168}{1050} - .25)/\sqrt{\frac{.25 \times .75}{1050}} = -6.09$. The $p$-value is calculated as $P(Z < -6.09) = \text{pnorm}(-6.09) \approx 0$ and since this is significantly less than .01, I would reject $H_o$ in favor of $H_a$.

3. $H_o : p = .035$ vs. $H_a : p < .035$. In this case, $\hat{p} = \frac{15}{500} = .03$. We calculate $z = (\hat{p} - p_o)/\sqrt{\frac{p_o(1-p_o)}{n}} = (\frac{15}{500} - .03)/\sqrt{\frac{.035 \times .965}{500}} = -6.08$. The $p$-value is calculated as $P(Z < -6.08) = \text{pnorm}(-6.08) = .271$. Since .271 > .05, we fail to reject $H_o$.

4. $H_o : p = 1/3$ vs. $H_a : p > 1/3$. In this case, $\hat{p} = \frac{346}{855} = .4$. We calculate $z = (\hat{p} - p_o)/\sqrt{\frac{p_o(1-p_o)}{n}} = (\frac{346}{855} - 1/3)/\sqrt{\frac{1/3 \times 2/3}{855}} = 4.13$. The $p$-value is calculated as $P(Z > 4.13) = 1 - \text{pnorm}(4.13) \approx 0$. So at the .01 level, I would reject $H_o$ in favor of $H_a$. And even though I rejected $H_o$ in favor of $H_a$, I’m still not that impressed with the tasters’ abilities to distinguish between the two types of wine. All this test shows it that they’re just slightly better then if they were just randomly choosing the better wine.

5. $H_o : p = .75$ vs. $H_a : p < .75$. In this case, $\hat{p} = .583$. We calculate $z = (\hat{p} - p_o)/\sqrt{\frac{p_o(1-p_o)}{n}} = (\frac{.583}{2/7} - .75)/\sqrt{\frac{.25 \times .25}{2/7}} = -3.27$. The $p$-value is calculated as $P(Z < -3.27) = \text{pnorm}(-3.27) \ll .01$, so at $\alpha = .01$, we would reject $H_o$ in favor of $H_a$. 

2