AP Statistics
Mr. Murphy
11.3 Large-Sample Inferences Concerning a Difference Between Two Proportions
HW Pg. 616 #11.41, 11.43, 11.55, 11.56

Goals: 1. Run a hypothesis test for the difference in means between two proportions.
2. Construct a confidence interval for the difference in means between two proportions.

Summary of Large-Sample z Tests for $p_1 - p_2 = 0$

Null hypothesis: $H_0: p_1 - p_2 = 0$

Test statistic: $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$$

Alternative hypothesis:	***P*-value:**
$H_a: p_1 - p_2 > 0$	Area under the z curve to the right of the computed z
$H_a: p_1 - p_2 < 0$	Area under the z curve to the left of the computed z
$H_a: p_1 - p_2 \neq 0$	(1) 2(area to the right of z) if z is positive or (2) 2(area to the left of z) if z is negative

DEFINITION

The combined estimate of the common population proportion is

$$\hat{p}_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{\text{total number of } S\text{'s in the two samples}}{\text{total of the two sample sizes}}$$

Ex1 Some people seem to believe that you can fix anything with duct tape. Even so, many were skeptical when researchers announced that duct tape may be a more effective and less painful alternative to liquid nitrogen, which doctors routinely use to freeze warts. A study was conducted at the Madigan Army Medical Center where patients with warts were randomly assigned to either the duct tape treatment or the more traditional freezing treatment. Those in the duct tape group wore duct tape over the wart for 6 days, then removed the tape, soaked the area in water, and used an emery board to scrape the area. This process was repeated for a maximum of 2 months or until the wart was gone. Data consistent with values in the study are summarized in the following table:

Treatment	n	Number with Wart Successfully Removed
Liquid nitrogen freezing	100	60
Duct Tape	104	88

Do the data suggest that freezing is less successful than duct tape in removing warts? Let the level of significance be 0.01.

A Large-Sample Confidence Interval for $p_1 - p_2$

When
1. the samples are *independently selected random samples* or *treatments were assigned at random to individuals or objects* (or vice versa), and
2. both *sample sizes are large:*

$n_1\hat{p}_1 \geq 10 \quad n_1(1-\hat{p}_1) \geq 10 \quad n_2\hat{p}_2 \geq 10 \quad n_2(1-\hat{p}_2) \geq 10$

a large-sample confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value})\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Ex2 Researchers at the National Cancer Institute released the results of a study examined the effect of weed-killing herbicides on house pets. Dogs, some of whom were from homes where the herbicide was used on a regular basis, were examined for the presence of malignant lymphoma. The following data was reported:

Group	Sample Size	Number with Lymphoma
Exposed	827	473
Unexposed	130	19

Use the given data to estimate the difference between the proportion of exposed dogs that develop lymphoma and the proportion of unexposed dogs that develop lymphoma.

Checkpoint
Multiple Choice

1. In a random sample of 200 University of Manitoba graduate students, it was found that 66% of them had previously attended some other college or university. In a random sample of 100 University of Waterloo graduate students, it was found that 35% of them had previously attended some other college or university. A 95% confidence interval for estimating the difference in proportions of graduate students who had previously attended some other college or university between the University of Manitoba and the University of Waterloo is:

(a) $(0.66-0.35)\pm 1.96\sqrt{(0.3366)(0.6633)\left(\frac{1}{200}+\frac{1}{100}\right)}$

(b) $(0.66-0.35)\pm 1.96\sqrt{\frac{(0.66)(0.34)}{200}+\frac{(0.35)(0.65)}{100}}$

(c) $(0.66-0.35)\pm 1.96\sqrt{(0.5566)(0.4433)\left(\frac{1}{200}+\frac{1}{100}\right)}$

(d) $(0.33-0.35)\pm 1.96\sqrt{(0.5566)(0.4433)\left(\frac{1}{200}+\frac{1}{100}\right)}$

(e) $(0.33-0.35)\pm 1.645\sqrt{(0.5566)(0.4433)\left(\frac{1}{200}+\frac{1}{100}\right)}$

The next two questions refer to the following situation:
One criticism of reforestation efforts after timber harvesting is that too few of the seedling survive. An experiment was conducted to assess if mulching the slash (limbs, roots, small branches, etc.) and leaving the mulch on the ground improves the survival rate compared to just leaving the slash on the ground. It is believed that mulching will cause the material to break down sooner and release the nutrients to the seedlings. A total of 500 seedlings were randomly assigned to the two treatments and the two year survival rate was measured. Of the 250 seedling receiving the “mulching” treatment, 75 survived; of the 250 seedlings receiving the “control” treatment, 55 survived.

2. The null and alternate hypotheses are: (m = mulch, c = control)

(a) $H_0: p_m = 0.22, H_a: p_m > 0.22$
(b) $H_0: \mu_m = 0.22,\ H_a: \mu_m > 0.22$
(c) $H_0: p_m - p_c = 0,\ H_a: p_m - p_c > 0$
(d) $H_0: \mu_m - \mu_c = 0,\ H_a: \mu_m - \mu_c > 0$
(e) $H_0: p_m - p_c = 0,\ H_a: p_m - p_c \neq 0$

3. The value of the test statistic and the p-value are:

(a) 2.76, 0.003
(b) 2.05, 0.042
(c) 2.76, 0.006
(d) 2.05, 0.021
(e) 2.05, 0.011

Free Response

1. Even though landlords participating in a telephone survey indicated that they would generally be willing to rent to persons with AIDS, it was wondered whether this was true in actual practice. To investigate, researchers independently selected two random samples of 80 advertisements for rooms for rent from newspaper advertisements in three large cities. An adult male caller responded to each ad in the first sample of 80 and inquired about the availability of the room and was told that the room was still available in 61 of these calls. The same caller also responded to each ad in the second sample. In these calls, the caller responded to each ad in the second sample. In these calls, the caller indicated that he was currently receiving some treatment for AIDS and was about to be released from the hospital and would require a place to live. The caller was told that a room was available in 32 of these calls. Based on this information, the study concluded that "reference to AIDS substantially decreased the likelihood of a room being described as available." Do the data support this conclusion? Carry out a hypothesis test with $\alpha = 0.01$.