Diff of Means, Var known

- Null hypothesis:

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

- Test statistic:

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Example

- Hermione hypothesizes that the tulips in her garden are, on average, 2 inches taller than the crocuses. By taking some measurements, she knows that her 28 tulips have an average height of 8 inches (and standard deviation 1.5), while the 33 crocuses have an average height of 5 inches (and standard deviation 1).

Questions I

- Test the hypothesis at the 5% level.
- What is the P-value?
- Give the 95% confidence interval for the difference in the two means.

Confidence Intervals

- The 100(1–α)% confidence interval on the difference of two means is:

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Required Sample Size (a)

- For the two-sided alternate hypothesis with significance level α, the sample size $n_1 = n_2 = n$ required to detect a true difference in means of Δ with power at least $1-\beta$ is:

$$n = \frac{(z_{\alpha/2} + z_{\beta})(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

Required Sample Size (b)

- If variances are known, and sample sizes are equal, the required sample size to guarantee an error no more than E at $100(1-\alpha)$ percent confidence is:

$$n = \left(\frac{z_{\alpha/2}}{E}\right)\left(\sigma_1^2 + \sigma_2^2\right)$$

Questions II

- Suppose that if the true difference in height is 3 inches, we want to detect this with probability at least 0.9. Find the required sample size.
- Suppose we limit the sample sizes of both kinds of flowers to be the same. What is the required sample size so that the error in estimating the difference in the means will be less than 0.5 inches with 95% confidence?

Diff of Means, Var unknown

- Null hypothesis:

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

$$\boxed{\sigma_1^2 = \sigma_2^2}$$

- Test statistic:

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{\underbrace{\boxed{n_1 - n_2 - 2}}_{\text{Degrees of freedom}}}$$

Diff of Means, Var unknown

- Null hypothesis:

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

$$\boxed{\sigma_1^2 \neq \sigma_2^2}$$

- Test statistic:

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \qquad \nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1+1} + \frac{\left(S_2^2/n_2\right)^2}{n_2+1}} - 2$$

Degrees of freedom

M&M Example

- Guido purchases packets of M&Ms on a regular basis from two locations: Easton and Allentown. He wonders if, on average, he gets the same number of M&Ms in each packet from the Easton store as he does from the Allentown store. Over the next few months, Guido carefully counts the number of M&Ms in each packet he buys.

M&M Example cont…

- He finds from the 10 packets he bought from the Easton store that there was an average of 28 M&Ms per packet with a sample standard deviation of 2.5. From the 12 packets he bought from the Allentown store, there was an average of 30 M&Ms with a sample standard deviation of 2.0.

M&M Example cont…

- Guido has read somewhere that he number of M&Ms in each packet has the same standard deviation. Would you say there is a difference in the Easton vs. Allentown M&M packets?
- What if the standard deviation were different? Have your conclusions changed?
- Construct a 95% confidence interval for the difference in M&M counts in each of the above two cases.

Paired t-test

- Null hypothesis:

$$H_0 : \mu_D = \Delta_0$$

- Test statistic:

$$T_0 = \frac{\overline{D} - \Delta_0}{S_D / \sqrt{n}}$$

Degrees of freedom = n–1

M&M Example Again

- Guido is a great fan of M&M chocolate, and has taken a special liking to the blue M&Ms. He wonders if there are as many blue M&Ms as there are red M&Ms in each packet. By sampling 12 packets of M&Ms, he finds that the difference in numbers between red and blue M&Ms has an average of 1.8 with a standard deviation of 1. Determine if there is any difference between the numbers.

Lehigh Mountain Hawk

- In November 1995, Lehigh University unveiled its new mascot, the Mountain Hawk. We want to compare the levels of support for the change from the "Engineer" to the "Mountain Hawk" for a mascot.

Mountain Hawk Questions

- Is there a difference in the support for the Mountain Hawk? Use α=0.05.
- What is the P-value for this test?
- Give a 95% confidence interval for the difference in the proportions.
- What is the power of the test?
- Determine the sample size needed to detect the difference w.p. at least 0.9.

Two Population Proportions

- Null hypothesis:

$$H_0 : p_1 = p_2$$

- Test statistic:

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad \hat{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

Info for Confidence Interval

$$p_1 - p_2 \sim N\left(\hat{p}_1 - \hat{p}_2, \sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}\right)$$

Calculating the β-error (2-sided)

$$\beta = \Phi\left(\frac{z_{\alpha/2}\sqrt{\overline{pq}\left(1/n_1 + 1/n_2\right)} - \left(p_1 - p_2\right)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right)$$

$$-\Phi\left(\frac{-z_{\alpha/2}\sqrt{\overline{pq}\left(1/n_1 + 1/n_2\right)} - \left(p_1 - p_2\right)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right)$$

$$\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{p_1\left(1-p_1\right)}{n_1} + \frac{p_2\left(1-p_2\right)}{n_2}}, \overline{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, \overline{q} = \frac{n_1\left(1-p_1\right) + n_2\left(1-p_2\right)}{n_1 + n_2}$$

Calculating the β-error (1-sided)

If $H_0 : p_1 > p_2$

$$\beta = \Phi\left(\frac{z_{\alpha/2}\sqrt{\overline{pq}\left(1/n_1 + 1/n_2\right)} - \left(p_1 - p_2\right)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right)$$

If $H_0 : p_1 < p_2$

$$\beta = 1 - \Phi\left(\frac{-z_{\alpha/2}\sqrt{\overline{pq}\left(1/n_1 + 1/n_2\right)} - \left(p_1 - p_2\right)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right)$$