1. An engineering education researcher wishes to determine whether there is a difference in the proportion of women enrolled in each of two engineering majors at U.S. colleges. She collects the following data from two random samples of U.S. college students enrolled in electrical and chemical engineering programs.

| **Engineering Major** | **Sample Size** | **Number of Women** | **Sample Proportion** |
|---|---|---|---|
| Electrical Engineering | 250 | 81 | $\hat{p}_E = \frac{81}{250} = 0.324$ |
| Chemical Engineering | 175 | 40 | $\hat{p}_C = \frac{40}{175} = 0.22857$ |

Construct a **90% confidence interval** for the difference in the proportion of women enrolled in electrical engineering versus chemical engineering.

(a) What are we estimating? (Define the parameters of interest.)

We are estimating $p_E - p_C$, where
$p_E$ = the true proportion of women enrolled in electrical engineering at U.S. colleges and
$p_C$ = the true proportion of women enrolled in chemical engineering at U.S. colleges.

(b) Verify that the necessary conditions for a confidence interval for the difference in two population proportions are met.

1. Independent random samples — it is stated that U.S. college students were randomly selected from each engineering program.
2. It can be reasonably assumed that $n_E = 250 < 5\%$ of all electrical engineering majors and $n_C = 175 < 5\%$ of all chemical engineering majors.
3. $n_E\hat{p}_E(1-\hat{p}_E) = 250(0.324)(1-0.324) = 54.756 > 10$ and $n_C\hat{p}_C(1-\hat{p}_C) = 175(0.22857)(1-0.22857) = 30.857 > 10$

(c) Compute the 90% confidence interval for the parameter defined in Part (a).

$$\begin{aligned}(\hat{p}_E - \hat{p}_C) \pm z_{.05}\sqrt{\frac{\hat{p}_E(1-\hat{p}_E)}{n_E} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}} &= (0.324 - 0.22857) \pm 1.645\sqrt{\frac{0.324(0.676)}{250} + \frac{0.22857(0.77143)}{175}} \\ &= 0.09543 \pm 0.07140 \\ &\approx (0.024, 0.167)\end{aligned}$$

(d) Interpret the confidence interval you found in Part (d).

We are 90% confident that the true difference in the proportion of women in electrical engineering and women in chemical engineering (E–C) is between 0.024 and 0.167.

(e) Based on the interval, can we infer that one type of engineering has a higher proportion of women enrolled?

Yes, we can infer that electrical engineering has between 2.4% and 16.7% more women enrolled than chemical engineering because zero is not in the interval and the interval contains all positive values.

2. Suppose you wish to investigate whether the proportion of Clemson undergraduates who drink coffee regularly is lower than the proportion of Clemson graduate students who drink coffee regularly. You gather the following data from randomly selected undergraduate and graduate students at Clemson University.

| **Student Type** | **Sample Size** | **Drink Coffee** | **Sample Proportion** |
|---|---|---|---|
| Undergraduate | 320 | 206 | $\hat{p}_U = \frac{206}{320} = 0.64375$ |
| Graduate | 350 | 238 | $\hat{p}_G = \frac{238}{350} = 0.68$ |

Conduct a **hypothesis test** for the difference in the two proportions at the $\alpha = 0.05$ level.

(a) Define the parameters of interest in context and state your hypotheses.

Let $p_U$ = the true proportion of Clemson undergraduates who drink coffee regularly
and $p_G$ = the true proportion of Clemson graduate students who drink coffee regularly.

$H_0 : p_U = p_G$
$H_1 : p_U < p_G$

(b) Check the appropriate conditions required for a valid hypothesis test.

1. Independent random samples — it is stated that Clemson undergraduate and graduate students were each randomly selected.
2. It can be reasonably assumed that $n_U = 320 < 5\%$ of all Clemson undergraduates and $n_G = 350 < 5\%$ of all Clemson graduate students.
3. $n_U\hat{p}_U(1 - \hat{p}_U) = 320(0.64375)(1 - 0.64375) = 73.3875 > 10$ and $n_G\hat{p}_G(1 - \hat{p}_G) = 350(0.68)(1 - 0.68) = 76.16 > 10$

(c) Compute the test statistic.

$$\hat{p}_U = 0.64375,\ \hat{p}_G = 0.68,\ \hat{p} = \frac{x_U + x_G}{n_U - n_G} = \frac{206 + 238}{320 + 350} = \frac{444}{670} = 0.66269$$

$$z_0 = \frac{\hat{p}_U - \hat{p}_G}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_U} + \frac{1}{n_G}\right)}} = \frac{0.64375 - 0.68}{\sqrt{0.66269(1 - 0.66269)\left(\frac{1}{320} + \frac{1}{350}\right)}} \approx -0.99$$

(d) Find the associated p-value for the hypothesis test.

p-value $= P(Z < -0.99) = 0.1611$

(e) State your conclusion in context of the problem.

Do not reject $H_0$ because p-value $= 0.1611 > \alpha = 0.05$. There is insufficient evidence to conclude that the proportion of Clemson undergraduates who drink coffee is lower than the proportion of Clemson graduate students who drink coffee.