6.10

(a) 61% is a sample statistic, it's the observed sample proportion.

(b) A 95% confidence interval can be calculated as follows:

$$\begin{aligned}\hat{p} \pm z^{\star}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.61 \pm 1.96\sqrt{\frac{0.61 \times (1-0.61)}{1578}} \\ &= 0.61 \pm 1.96 \times 0.012 \\ &= 0.61 \pm 0.024 \\ &(0.586, 0.634)\end{aligned}$$

We are 95% confident that approximately 58.6% to 63.4% of Americans think marijuana should be legalized.

(c) 1. Independence: The sample is random, and comprises less than 10% of the American population, therefore we can assume that the individuals in this sample are independent of each other

2. Success-failure: The number of successes (people who said marijuana should be legalized: $1578 \times 0.61 = 962.58$) and failures (people who said it shouldn't be: $1578 \times 0.39 = 615.42$) are both greater than 10, therefore the success-failure condition is met as well.

Therefore the distribution of the sample proportion is expected to be approximately normal.

(d) Yes, the interval is above 50%, suggesting, with 95% confidence, that the true population proportion of Americans who think marijuana should be legalized is greater than 50%.

6.13

(a) The hypotheses are as follows:

$H_0 : p = 0.5$ (Results are equivalent to randomly guessing)

$H_A : p \neq 0.5$ (Results are different than just randomly guessing)

Before conducting the hypothesis test, we must first check that the conditions for inference are satisfied.

1. Independence: The sample is random, therefore whether or not one person in the sample can identify a soda correctly in independent of another.
2. Success-failure: $80 \times 0.5 = 40 > 10$ and $80 \times 0.5 = 40 > 10$. Since the observations are independent and the success-failure condition is met, $\hat{p}$ is expected to be approximately normal.

The test statistic and the p-value can be calculated as follows:

$$\hat{p} = \frac{53}{80} = 0.6625$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.6625 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{80}}} = \frac{0.1625}{0.0559} = 2.91$$

$$p - value = 2 \times P(\hat{p} > 0.6625 \mid p = 0.5) = 2 \times P(Z > 2.91) = 2 \times (1 - 0.9982) = 0.0036$$

Since the p-value $< \alpha$ (use $\alpha = 0.05$ since not given), we reject the null hypothesis. Since we rejected H_0 and the point estimate suggests people are better than random guessing, we can conclude the rate of correctly identifying a soda for these people is significantly better than just by random guessing.

(b) If in fact people cannot tell the difference between diet and regular soda and they were randomly guessing, the probability of getting a random sample of 80 people where 53 or more identify a soda correctly (or 53 or more identify a soda incorrectly) would be 0.0036.

6.24

(a) The hypotheses are:

$$H_0 : p_{CA} = p_{OR}$$
$$H_A : p_{CA} \neq p_{OR}$$

We have confirmed in Exercise ?? that the independence condition is satisfied but we need to recheck the success-failure condition using $\hat{p}_{pool}$ and expected counts.

$$success_{CA} = n_{CA} \times p_{CA} = 11,545 \times 0.08 = 923.6 \approx 924$$
$$success_{OR} = n_{OR} \times p_{OR} = 4,691 \times 0.088 = 412.8 \approx 413$$
$$\hat{p}_{pool} = \frac{success_{CA} + success_{OR}}{n_{CA} + n_{OR}} = \frac{924 + 413}{11,545 + 4,691} = \frac{1,337}{16,236} \approx 0.0821 - \hat{p}_{pool} \quad = 1 - 0.082 = 0.918$$

$$11,545 \times 0.082 = 946.69 > 10 \qquad 11,545 \times 0.918 = 10598.31 > 10$$
$$4,691 \times 0.082 = 384.662 > 10 \qquad 4,691 \times 0.918 = 4306.338 > 10$$

Since the observations are independent and the success-failure condition is met, $\hat{p}_{CA} - \hat{p}_{OR}$ is expected to be approximately normal. Next we calculate the test statistic and the p-value:

$$\begin{aligned} Z &= \frac{(\hat{p}_{CA} - \hat{p}_{OR}) - (p_{CA} - p_{OR})}{\sqrt{\frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_{CA}} + \frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_{OR}}}} \\ &= \frac{(0.08 - 0.088) - 0}{\sqrt{\frac{0.082 \times 0.918}{11,545} + \frac{0.082 \times 0.918}{4,691}}} \\ &= \frac{-0.008}{0.00475} = -1.68 \end{aligned}$$

$$p - value = P(|\hat{p}_{CA} - \hat{p}_{OR}| > 0.008 \mid (p_{CA} - p_{OR}) = 0) = 2 \times P(|Z| > 1.68) = 2 \times 0.0465 = 0.093$$

Since the p-value $> \alpha$ (use $\alpha = 0.05$ since not given), we fail to reject H_0 and conclude that the data do not provide strong evidence that the rate of sleep deprivation is different for the two states.

(b) Type II, since we may have incorrectly failed to reject H_0.

6.25

(a) College grads: $\frac{154}{438} = 0.352$
Non-college grads: $\frac{132}{389} = 0.339$

(b) Let p_{CG} represent the proportion of college graduates who support offshore drilling, and p_{NCG} represent the proportion of non-college graduates who do so. Then,

$$H_0 : p_{CG} = p_{NCG}$$
$$H_A : p_{CG} \neq p_{NCG}$$

Before calculating the test statistic we should check that the conditions are satisfied.

1. Independence: Both samples are random and unrelated, so independence is satisfied.
2. Success-failure: First we need to find $\hat{p}_{pool}$ and then use that to calculate the numbers of expected successes and failures in each group.

$$\hat{p}_{pool} = \frac{success_{CG} + success_{NCG}}{n_{CG} + n_{NCG}} = \frac{154 + 132}{438 + 389} = \frac{286}{827} = 0.346$$
$$1 - \hat{p}_{pool} = 1 - 0.346 = 0.654$$

$$438 \times 0.346 = 151.548 > 10 \qquad 438 \times 0.654 = 286.452 > 10$$
$$389 \times 0.346 = 134.594 > 10 \qquad 389 \times 0.654 = 254.406 > 10$$

Since the observations are independent and the success-failure condition is met, $\hat{p}_{CG} - \hat{p}_{NCG}$ is expected to be approximately normal. Next we calculate the test statistic and the p-value:

$$Z = \frac{(\hat{p}_{CG} - \hat{p}_{NCG}) - 0}{\sqrt{\frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_{CG}} + \frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_{NCG}}}}$$
$$= \frac{(0.352 - 0.339)}{\sqrt{\frac{0.346 \times 0.654}{438} + \frac{0.346 \times 0.654}{389}}} = \frac{0.013}{0.033} = 0.39$$

$$p - value = P(|\hat{p}_{CG} - \hat{p}_{NCG}| > 0.013 \mid (p_{CG} - p_{NCG}) = 0) = P(|Z| > 0.39) = 2 \times 0.3483 = 0.6966$$

Since the p-value $> \alpha$ (0.05), we fail to reject H_0. The data do not provide strong evidence of a difference between the proportions of college graduates and non-college graduates who support off-shore drilling in California.