1. $\bar{X} \sim \mathcal{N}\left(10, \frac{20^2}{25}\right)$, approximately so the answer is (e).

2. (d)

3. Both histograms and boxplots indicate a right skewed distribution for the population. However the histogram for the sample means is close to a bell-shaped curve. ($\bar{X} \simeq \mathcal{N}(1, (0.1)^2)$.)

4. $P = 0.98$ indicates that the observed test statistic is consistent with the null hypothesis. A statistical test does not prove that H_0 is true or false.

5. Test $H_0 : p = 0.6$ versus $H_1 : p > 0.6$.

(a) There were 8 successes in 10 trials. The test statistic is X, the number of successes. If H_0 is true then $X \sim B(10, 0.6)$. Large values of X argue for H_1. Thus the P-value is

$$P = P(X \geq 8) = 1 - P(X \leq 7) = 0.1673.$$

The data are consistent with H_0.

(b) If we observe 50 trials then $X \sim \mathcal{B}(50, p)$. If the observed number of successes is 40 then the P-value is

$$P = \mathrm{P}(X \geq 40) = 0.0022, \quad \text{using R and 1-pbinom(39,50,0.6).}$$

Alternatively using the approximating normal $Y \sim \mathcal{N}(30, 12)$ we get

$$P \simeq \mathrm{P}\left(Z \geq \frac{40 - 0.5 - 30}{\sqrt{12}}\right) = 0.0031.$$

In this case there is strong evidence to argue against $H_0 : p = 0.6$.

(c) In both (a) and (b) the observed proportion is 0.8. The observed proportion is not the only factor to be considered. The sample size is critical. The more observations we have the better our understanding of the population of interest.

6. Let X denote the number developing lung cancer, $X \sim \mathcal{B}(90, p)$. Test $H_0 : p = 0.10$ against the alternative $H_1 : p > 0.10$. Large values of X argue for H_1. We observe $x = 19$. The P-value is

$$P = \mathrm{P}(X \geq 19) = 1 - \mathrm{P}(X \leq 18) = 0.0013, \quad \text{from R.}$$

(Using the normal approximation we get $Y \sim N(9, 8.1)$ so

$$P \simeq \mathrm{P}\left(Z \geq \frac{19 - 0.5 - 9}{\sqrt{8.1}}\right) = \mathrm{P}(Z \geq 3.338) = 0.0004.$$

Thus the outcome of 19 deaths from lung cancer is extremely unlikely if the incidence rate is 10%. There is strong evidence to argue for H_1.

The following functions in R can be used for the quick calculation of tests for proportions: `binom.test()` and `prop.test()`. Check lecture notes for some worked through examples.