1. The claim is made that at least 41% of the U.S. voting population approves of President Obama’s job performance. State the null and alternative hypotheses.

**null hypothesis** $H_0$: support $\geq 0.41$

**alternative hypothesis** $H_a$: support $< 0.41$

2. The claim is made that the percentage of the population with Type O blood exceeds 30%. What would a type I error be in this case?

The claim is an alternative hypothesis because there is no equality.

The corresponding null hypothesis: Type O blood prevalence $\leq 30$

A type I error occurs when a true null hypothesis is rejected:

We conclude that Type O blood prevalence exceeds 30% when in fact the prevalence is at most 30%.

3. A scatter plot contains a sample of 15 data points, for which the Pearson correlation coefficient is calculated to be $r = 0.605$. At level of significance $\alpha = 0.05$, can I make the claim that the bigger set of data has a linear correlation?

The critical value of $r$ for this level of significance is 0.514.

At this level of significance, we can make the claim that the bigger set of data has a linear correlation.

4. A set of data is normally distributed with standard deviation $\sigma = 1$. For the null hypothesis $\mu = 0$, sketch the rejection region corresponding to level of significance $\alpha = 0.02$.

*rejection region is shaded red*
5. A large state university conducts a survey to calculate the mean annual income of their graduates. The university surveyed 500 graduates, and found the mean income of the surveyed group to be $\bar{x} = $45,000 with a sample standard deviation of $s = $15,000.

(a) Construct a 95%-confidence interval for the mean income of this university’s graduates.

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} = \frac{15000}{\sqrt{500}} \approx 671
\]

\[
E = z\sigma_{\bar{x}} \approx (1.96)(671) \approx 1315
\]

The 95% confidence interval is $(15000 - 1315, 15000 + 1315) = (13685, 16315)$.

(b) The claim is made that the mean income of this university’s graduates exceeds $43,000. With level of significance $\alpha = 0.01$, can this claim be supported?

**null hypothesis** $H_0$: mean income $\leq 43000$

**rejection region**: income $\geq$ mean$+z\sigma_{\bar{x}}$

income $\geq 43000 + (2.33)(1315) = 46064$

The observed mean of 45000 is not in the rejection region.

The null hypothesis cannot be rejected; our data cannot support the claim that the mean income exceeds $43000 to the desired level of significance.

6. A representative sample of 400 people finds that 60 are suffering from a cold. Construct a 99% confidence interval for the proportion of the population which has a cold.

\[
\hat{p} = \frac{60}{400} = 0.15 \quad \text{and} \quad \hat{q} = 0.85
\]

\[
\sigma_\hat{p} = \sqrt{\frac{\hat{p}\hat{q}}{n}} \approx \sqrt{\frac{\hat{p}\hat{q}}{400}} = \sqrt{\frac{(0.15)(0.85)}{400}} \approx 0.018
\]

\[
E = z\sigma_\hat{p} \approx (2.33)(0.018) \approx 0.04
\]

The 99% confidence interval is $(0.15 - 0.04, 0.15 + 0.04) = (0.11, 0.19)$. 
Two large independent samples of cancer patients are studied to see if a healthy diet improves survival times. The 100 patients on the healthy diet had a mean survival time of \( \bar{x}_1 = 5.5 \) years, with sample deviation \( s_1 = 3 \) years, while the other 100 patients on regular diets had a mean survival time of \( \bar{x}_2 = 4.5 \) years, with a sample deviation of \( s_2 = 2 \) years. Can the claim be made that the healthy diet improves survival times, with level of significance \( \alpha = 0.05 \) ?

**original (and alternative) hypothesis** \( H_a: \mu_1 > \mu_2 \) or \( \mu_1 - \mu_2 > 0 \)

**null hypothesis** \( H_0: \mu_1 - \mu_2 \leq 0 \)

\[
\sigma_{\bar{x}_1 - \bar{x}_2} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{3^2}{100} + \frac{2^2}{100}} \approx 0.36
\]

**standardized test statistic:** 
\[
z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} \approx \frac{(5.5 - 4.5) - 0}{0.36} \approx 2.78
\]

**rejection region:** \( z > 1.64 \)

The standardized test statistic lies in the rejection region

Reject the null hypothesis

The claim is supported that a healthy diet improves survival times.
8. For the paired data above, decide whether the (Pearson) correlation coefficient $r$ between $x$ and $y$ values is
(a) close to 1  (b) close to 0  (c) close to $-1$
It is not necessary to calculate the coefficient $r$ exactly!

$r$ is close to 1.

9. For the data plotted above, draw (visually) a best-fit line. Then write down an equation for the best-fit line you have drawn.
It is not necessary to calculate the least-squares best-fit line!

Line contains $(10, 17)$ and $(22, 40)$

$$m = \frac{40 - 17}{22 - 10} = \frac{23}{12} \approx 1.9$$

$$y - 17 \approx 1.9(x - 10)$$

$$y \approx 1.9x - 2$$