(1) **Comparing Two Populations 1**

A study of the differences in cognitive function between normal individuals and patients diagnosed with schizophrenia was published in the American Journal of Psychiatry (Apr. 2010). The total time (in minutes) a subject spent on the Trail Making Test (a standard psychological test) was used as a measure of cognitive function. The researchers theorize that the mean time on the Trail Making Test for schizophrenics will be larger than the corresponding mean for normal subjects. The data for independent random samples of 41 schizophrenics and 49 normal individuals yielded the following results:

	Schizophrenia	Normal
Sample size	41	49
Mean time	104.23	62.24
Standard deviation	62.24	16.34

(a) Define the parameter of interest to the researchers.

(b) Set up the null and alternative hypothesis for testing the researchers' theory.

(c) The researchers conducted the test, part (b), and reported a p-value of .001. What conclusions can you draw from this result? (Use $\alpha = 0.01$)

(d) Find a 99% confidence interval for the target parameter. Interpret the result. Does your conclusion agree with that of the previous part?

a) We have large sample sizes $n_1 = 41 \geq 30$ and $n_2 = 49 \geq 30$, hence we have approximately $\overline{X_1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\overline{X_2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$. From this we obtain $\overline{X_1} - \overline{X_2} \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

The target parameter is $\mu_1 - \mu_2$. We take the idea from slide 32 of lecture 11.

b) We would choose $H_0: \mu_1 - \mu_0 = 0$ vs. $H_1: \mu_1 - \mu_0 > 0$

The p-value is

$$\mathbb{P}\left(\overline{X_1} - \overline{X_2} \geq \bar{x}_1 - \bar{x}_2\right) = 1 - \mathbb{P}\left(\frac{\overline{X_1} - \overline{X_2}}{s} < \frac{\bar{x}_1 - \bar{x}_2}{s}\right) = 1 - \phi\left(\frac{\bar{x}_1 - \bar{x}_2}{s}\right) \approx 10^{-5} ?\,?$$

c) Since the p-value is very small, we reject H_0. If we decide to use a confidence level of $\alpha = 1/100$, then we would reject H_0.

$$s := \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

d) We would like to find $d \in \mathbb{R}$, such that

$$1-\alpha = \mathbb{P}\left(\bar{x}_1 - \bar{x}_2 - d \leq \overline{X_1} - \overline{X_2} < \bar{x}_1 - \bar{x}_2 + d\right) = \mathbb{P}\left(-d \leq \overline{X_1} - \overline{X_2} - (\bar{x}_1 - \bar{x}_2) < d\right)$$

$$= \mathbb{P}\left(\frac{-d}{s} \leq \frac{\overline{X_1} - \overline{X_2} - (\bar{x}_1 - \bar{x}_2)}{s} < \frac{d}{s}\right) = 1 - 2\,\phi\left(\frac{-d}{s}\right) \Leftrightarrow \frac{-d}{s} = \Phi^{-1}\left(\frac{\alpha}{2}\right) \Leftrightarrow d = -\Phi^{-1}\left(\frac{\alpha}{2}\right)s$$

The interval is given by $[\bar{x}_1 - \bar{x}_2 - d, \bar{x}_1 - \bar{x}_2 + d) \approx [16.24, 67.74)$

The interval does not contain 0, hence we are 99% confident, that the hypothesis H_0 can not be rejected.