hw_lect18_1 use z or t.

For the data you collected, consider one of the continuous variables (call it y), and one of the categorical variables (call it x). Let mu1 denote the true mean of y when x = (first lelvel of x), and mu2 denote the true mean of y when x= (2nd level of x).

a) compute a t-based, 2-sided, 95% C.I. for mu1-mu2.

b) Is there evidence from data that mu1 and mu2 are diffiererent?

Suppose y contains the cont. data, and x takes values 1,2.

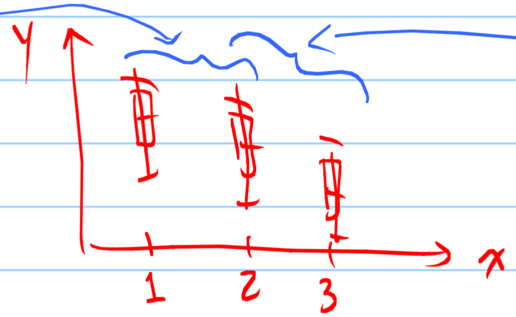
Generally, your code would look like this:

```
t.test( y[x1 == 1] , y[x2 == 2] )
```

If you want to compare the y at 2 other levels of x, e.g. 2,3:

```
t.test( y[x1 == 2] , y[x2 == 3] )
```

In a previous hw problem you made a plot like this:



The above t.tests allow you to test whether there is a difference in the true means of y at 2 different levels of x.

Later, in Ch.9 we will learn about a more efficient way of testing the means of y at different x's.

hw_lect18_2

Let pi_1 denote the true proportion of defective bridges in the USA, and pi_2 .... in Canada. A sample of n1=80, and n2=50 bridges from the two countries, respectively, is taken, and it is found that 21% of the bridges in the USA, and 10% of the bridges in Canada are defective. At 95% confidence level

a) Is there evidence that the true proportions are different?

b) Is there evidence that pi_1 is larger than pi_2? This type of question requires a 1sided CI which we are skipping this quarter. So, skip part b.

a) 2-sided CI : $(p_1 - p_2) \pm 1.96\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

$(0.21 - 0.10) \pm 1.96\sqrt{\frac{\cdots}{80} + \frac{\cdots}{50}}$

$0.11 \pm 1.96(0.0622) = 0.11 \pm 0.1219$

Does include zero. $= (-0.0119, 0.232)$

$\therefore$ There is no evidence that $\pi_1 \neq \pi_2$

$\therefore$ We cannot conclude that $\pi_1 \neq \pi_2$ (nor $\pi_1 = \pi_2$).

Part b) is only FYI

b) $\pi_1 \overset{?}{>} \pi_2 \Rightarrow \pi_1 - \pi_2 \overset{?}{>} 0 \Rightarrow$ Lower Conf. bound for $\pi_1 - \pi_2$:

$(p_1 - p_2) - 1.645\sqrt{\cdots + \cdots} = 0.11 - 1.645(0.0622) = 0.007$

$\therefore$ 95% Conf. That $\pi_1 - \pi_2 > .007$.

This interval does NOT include zero.

$\therefore$ There is evidence that $\pi_1 > \pi_2$.

Note the "apparent" contradiction in the conclusions!

Same Data! Different conclusions (from different questions) hw_lect19_1 use t- distr.

```
# Consider the following data on x1 and x2 which was collected in a paired design:
x1 = c(-0.27, -0.14,  1.61,  0.09,  0.00,  2.07,  0.56, -1.67, -0.51, -0.54)
x2 = c(-0.32,  0.20,  1.93,  0.54,  0.75,  1.77,  0.84, -0.29, -0.33,  0.17)
# a) Compute a 2-sided, 95% CI for the difference between the two true means. You may use R to do
# simple claculations, but use the CI formulas derived in class.

n = length(x1)
d = x1 - x2
dbar = mean(d)
s_d = sd(d)
dbar - 2.262 *s_d/sqrt(n)    #  -0.738  (t*=2.262 is from table 4, 2-sided 95% column with df=10-1 =9; in R, it can be obtained this way: qt(.05/2, df = 10-1, lower.tail=F))
dbar + 2.262 *s_d/sqrt(n)    #  -0.073

# b) Provide one interpretation of the observed CI, AND state the conclusion in English, i.e., the "corollary."

# We are 95% confident that mu1-mu2 is in the interval (-0.7380, -0.073)
# Corollary: There is a difference between the two means

# BTW, you can "test" that x1 and x2 are paired by looking at their scatterplot:

plot(x1,x2)          # I see a linear association.

# c) Consider the following data, which is the same as above, except the cases in x2 have been randomly shuffled. Compute an
#  appropriate 95% 2-sided CI.
y1 = c(-0.27, -0.14,  1.61,  0.09,  0.00,  2.07,  0.56, -1.67, -0.51, -0.54)
y2 = c( 0.20 , 0.54, -0.33,  1.93, -0.32,  1.77,  0.75,  0.17, -0.29,  0.84)

n1 = length(y1)
n2 = length(y2)
y1bar = mean(y1)
y2bar = mean(y2)
sd1 = sd(y1)
sd2 = sd(y2)
std.err = sqrt(sd1^2/n1 + sd2^2/n2)
df_welch = (sd1^2/n1 + sd2^2/n2)^2/( (sd1^2/n1)^2/(n1-1) + (sd2^2/n2)^2/(n2-1) ) # 16.76825 i.e., about 17
(y1bar - y2bar) - 2.110 *std.err        # -1.309      t* = 2.110 from Table IV or from R:  qt(.05/2, df = 17, lower.tail=F)
(y1bar - y2bar) + 2.110*std.err         #  +0.498

# d) Provide one interpretation of the observed CI, AND state the conclusion in English, i.e., the "corollary."

# We are 95% confident that mu1 - mu2 is in the interval (-1.309, +0.498)
# Corollary: We cannot tell if there is a difference between the two means.

# e) Which one is narrower?
# The width of the paired interval is
  2.262 *s_d/sqrt(n)    # 0.3326533
# The width of the unpaired interval is
  2.110 *std.err        # 0.9035258
# The paired one is narrower, i.e., more precise (or more reliable). And that is why the paired  CI is able to
# "see" a difference between the means while the unpaired CI cannot detect the difference.
```