Solutions to Recitation 10-27-09

Announcement: Equations 3, 4, 5 will be provided on the exam (% CI for $\mu_x - \mu_y$, $p_x - p_y$, and the improved % CI for μ using correlation). The equation for stratified CI will also be provided (this will be the only new equation next week).

Q1. point estimate for $\mu_y = \bar{y}$
$= \text{sample mean of } \bar{y}$
$= 12.22$

Q2. improved point estimator for $\mu_y = \bar{y} + (\mu_x - \bar{x}) R \frac{s_y}{s_x}$
$= 12.22 + (17.3 - 16.8)\, 0.64 \frac{2.73}{1.1}$
$= 13.014$

* this is the same equation as the one in lecture, we just pulled a negative sign from $(\bar{x} - \mu_x)$ and placed it outside the brackets.

R = correlation between X, y

Q3. Find the z-based 95% CI for μ_y.

$\bar{y} \pm z_{\alpha/2} \frac{s_y}{\sqrt{n}}$

$12.22 \pm 1.96 \frac{2.73}{\sqrt{20}}$

12.22 ± 1.196

* if the question asked for the t-based 95% CI for μ_y, we use:

$\bar{y} \pm t_{\alpha/2} \frac{s_y}{\sqrt{n}}$

where $t_{\alpha/2} = 2.093$ since $df = n-1 = 20-1 = 19$

* the question will always state whether to find the z-based or t-based CI

* if the question ~~asks~~ was without-replacement, we use:

$\bar{y} \pm \left(z_{\alpha/2} \frac{s_y}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}\right)$

where N = population size and N will be stated in the question

Q4. Find the z-based 95% CI for μ_y.

$[\bar{y} + (\mu_x - \bar{x}) R \frac{s_y}{s_x}] \pm z_{\alpha/2} \frac{s_y}{\sqrt{n}} \sqrt{1-R^2}$

$13.014 \pm 1.96 \frac{2.73}{\sqrt{20}} \sqrt{1-0.64^2}$

13.014 ± 0.919

Q5. The improved estimator for μ_y is larger than $\bar{y}$ because μ_x is greater than $\bar{x}$ ($\mu_x - \bar{x}$ is positive).

Q6. The ME for the improved CI for μ_y is 0.919, compared to the ME for the normal CI for μ_y which is 1.196. Since the improved CI has a smaller ME, it is also the smaller CI of the two, and the smaller the CI is, the more precise (better) it is.

*Note: When we calculate the improved CI for μ_y, we have paired data. Paired data means we take ~~the same~~ measurements on the same object but at different times.

eg.) we measure test scores on the same 20 student but once for exam 1 and once again for exam 2.

Paired data is ~~do~~ <u>not</u> independent because we take measurements on the same object.

Q7 Find a z-based 95% CI for $\mu_x - \mu_y$.

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

$$17.4 - 14.8 \pm 1.96\sqrt{\frac{1.9^2}{30} + \frac{3.8^2}{40}}$$

$$2.6 \pm 1.360$$

Q8. The idea is that after we have calculated the CI in Q7, how can we find a even better CI using s_x and s_y. We can use s_x and s_y to find the optimal sample sizes for X and y (n_x and n_y). Optimal sample sizes will in turn give us the narrowest CI (CI with smallest ME). The idea is, ~~to~~ we have a total of 70 samples (30+40) and we want to redistribute among n_x and n_y.

optimal proportion for n_x $= \frac{s_x}{s_x + s_y} = \frac{1.9}{1.9+3.8} = \frac{1}{3}$

optimal n_x $= \frac{1}{3} \times 70 = 23$ (multiple by total # of samples)

optimal proportion for n_y $= \frac{s_y}{s_x + s_y} = \frac{2}{3}$ (or $= 1 - \frac{1}{3}$)

optimal n_y $= \frac{2}{3} \times 70 = 47$

* sample sizes need to be whole numbers so we round to nearest 1

Out[23]//MatrixForm=

Q9

H	H	H	T	H	H	H	H	T	T	T	T	H	T	T	H	H	T	H	H	H	H	H	H	H	H	T	T	T	H	H	H	H	H	T	H	H	T	H
T	H	H	H	H	T	H	T	H	T	H	T	T	H	H	T	H	H	T	T	H	H	H	T	T	T	H	T	H	T	H	T	T	H	T	T	T	T	H
T	H	H	H	H	H	T	H	H	H	H	H	T	H	H	H	T	T	T	H	H	H	H	T	H	H	T	T	H	T	T	T	H	T	H	H	H	H	H
H	T	H	T	H	H	T	T	T	H	H	H	T	H	H	T	H	T	T	T	T	H	H	T	H	T	T	T	H	H	H	T	H	H	H	H	H	T	T
T	T	T	T	T	H	T	H	H	T	H	H	H	T	T	T	H	H	T	T	H	H	H	T	T	T	T	T	H	T	H	T	H	H	H	T	T	H	H
H	T	T	T	T	H	T	T	H	T	H	T	H	H	H	T	H	T	T	H	T	T	H	H	H	H	T	T	T	H	T	H	H	T	T	H	H	H	H
T	T	H	T	T	T	H	H	H	T	T	H	H	H	T	H	T	T	H	H	T	T	H	H	T	H	T	H	H	H	H	T	H	T	T	H	H	T	H
H	T	H	T	T	H	H	T	T	H	H	H	H	H	T	T	T	T	T	H	T	H	H	H	H	H	T	T	H	H	H	T	H	T	H	H	H	H	T
T	T	T	H	T	T	T	T	T	T	T	T	H	H	H	T	T	H	H	H	H	H	H	T	H	H	H	T	T	H	T	H	T	T	H	T	T	T	T
T	T	H	H	H	T	H	H	H	T	H	H	T	T	H	H	T	T	T	T	T	T	H	H	H	H	T	H	T	H	T	T	H	H	H	T	T	H	T
H	H	T	H	H	H	T	H	T	H	T	H	H	T	T	H	T	H	H	T	T	T	H	T	T	H	T	H	H	T	H	H	T	H	T	H	T	H	H
T	T	H	T	T	H	H	H	T	T	H	T	T	T	H	H	H	T	T	H	H	H	H	H	H	H	T	H	H	T	T	H	T	H	H	H	T	T	H
H	H	H	T	T	H	T	T	T	H	T	H	H	H	H	T	H	H	H	H	T	T	T	H	H	T	H	H	H	T	T	T	H	T	T	T	T	T	H
T	H	H	H	H	T	H	H	H	H	H	H	T	T	H	T	T	T	H	T	H	H	H	H	H	T	H	T	H	H	T	H	T	H	T	H	T	H	T
H	H	H	T	H	T	H	H	T	T	T	H	T	T	H	T	T	H	H	T	H	T	T	H	H	H	T	H	T	H	T	T	H	H	H	H	H	T	H
H	H	T	T	T	T	T	H	H	T	H	H	H	T	H	H	H	H	T	T	T	H	T	H	T	T	H	H	H	H	T	T	T	T	T	H	T	H	T
H	H	T	T	H	H	T	T	H	H	T	T	H	H	T	H	T	T	H	T	H	T	T	H	H	H	T	H	T	T	T	H	H	H	H	T	T	H	H
H	H	H	H	H	T	H	H	T	T	H	H	H	H	H	T	T	H	H	T	T	T	H	T	H	H	H	H	H	T	H	T	T	H	H	H	T	T	H
H	H	T	T	T	T	H	T	H	H	T	T	H	H	H	H	H	H	T	T	T	H	T	H	H	H	H	H	T	H	H	T	T	H	T	H	T	T	T
H	H	H	T	T	H	H	T	T	T	H	H	H	H	T	T	T	H	T	H	T	H	H	T	H	H	T	H	T	T	T	H	H	T	H	T	H	T	T

In[29]:= **HHrun = {2, 3, 3, 6, 9, 14, 8, 7, 14, 4, 2, 7, 2, 3, 2, 2, 2, 2, 2, 2}**

Out[24]//MatrixForm=

Q10

T	H	T	T	T	H	H	T	H	T	H	T	T	T	H	T	H	T	H	H	H	T	T	H	H	T	H	H	T	H	H	T	T	T	T	H	T	H	H
T	T	H	T	H	H	T	H	T	H	T	H	T	T	H	T	T	T	H	H	T	H	T	H	H	H	T	T	H	T	H	H	H	H	T	T	T	T	T
T	T	H	T	H	T	T	T	H	T	T	T	T	H	T	H	H	T	T	T	H	T	H	T	H	T	T	T	T	H	H	T	T	H	H	T	T	H	H
T	T	T	T	T	T	T	T	T	H	H	H	T	T	T	H	H	H	T	T	T	H	T	H	T	T	T	T	H	H	T	H	T	H	T	T	T	H	T
T	H	T	H	T	H	T	H	T	H	H	T	H	H	T	T	H	T	T	H	T	H	H	H	T	T	H	H	T	T	H	H	T	H	T	T	T	T	H
T	T	H	H	H	T	T	T	H	T	H	H	T	T	H	T	H	H	T	H	T	H	T	H	T	H	T	T	H	H	H	T	H	T	T	H	T	T	H
T	H	H	T	H	H	T	H	H	T	H	H	H	H	H	H	H	T	H	T	H	T	T	T	T	T	T	H	T	H	T	H	H	H	H	H	T	H	H
T	T	T	H	T	T	H	T	H	T	T	T	T	T	H	T	H	H	H	T	T	T	T	T	T	H	H	T	T	T	H	T	T	T	T	H	T	T	T
H	T	T	H	T	H	H	H	T	H	H	T	H	H	T	T	H	H	T	H	H	T	T	T	T	H	T	T	T	T	H	H	T	H	T	T	H	H	T
T	T	H	H	T	H	H	H	T	H	T	H	H	H	T	T	H	T	H	H	H	T	T	T	H	T	H	H	T	H	H	T	T	H	H	H	H	T	H
T	T	T	H	T	T	H	T	T	T	H	H	T	T	H	T	H	T	T	T	T	H	H	T	H	T	H	T	H	H	H	H	H	H	H	H	H	H	T
T	H	T	H	H	H	H	T	T	H	H	T	H	H	T	H	T	T	T	T	T	H	T	H	T	T	H	H	T	T	T	H	T	T	T	H	T	H	T
T	T	H	H	T	H	H	T	H	H	H	T	H	H	H	H	H	T	T	T	H	H	T	T	H	T	H	H	T	H	H	T	H	T	H	T	H	T	T
T	H	H	T	T	T	H	H	H	H	H	H	T	T	H	T	T	H	T	H	H	T	H	H	H	T	T	T	H	T	T	H	H	T	T	H	H	H	[illegible]
H	T	T	T	H	T	H	T	H	T	T	T	H	T	H	T	T	H	T	T	T	H	H	T	H	H	H	H	T	T	H	T	H	T	H	H	T	H	T
H	H	H	H	T	H	T	T	T	H	H	T	H	H	H	H	T	H	H	T	H	H	H	H	H	T	H	T	H	H	T	T	T	H	T	T	T	H	T
H	H	T	T	H	H	H	T	T	H	T	H	T	T	T	H	T	T	H	T	T	H	T	T	T	T	T	H	H	H	H	H	H	H	H	T	H	H	H
H	T	H	T	T	T	T	H	H	T	H	H	T	H	H	T	T	T	T	H	H	T	H	H	H	H	H	H	H	H	T	T	H	H	T	H	H	H	H
T	T	T	T	T	T	T	T	H	H	H	H	T	T	H	H	T	T	H	H	H	T	T	H	T	T	H	T	T	H	H	H	T	H	T	T	T	H	T
T	T	H	T	T	T	H	H	T	T	H	T	T	T	H	H	H	T	T	H	T	T	T	T	H	H	T	T	H	T	T	T	H	T	T	T	T	H	H
H	H	H	T	H	H	T	T	T	H	T	T	T	H	H	T	H	H	T	H	H	T	T	T	T	H	H	T	T	H	H	T	T	H	T	H	H	T	H
H	H	T	T	H	H	H	T	H	T	T	T	H	H	H	T	H	H	H	H	H	T	T	H	T	H	H	T	T	H	T	H	H	T	T	H	H	H	H
T	T	H	H	T	T	H	T	T	T	T	H	H	T	H	T	H	H	H	T	T	T	H	H	H	T	T	T	H	T	T	T	H	H	T	T	T	H	[illegible]
T	H	H	T	T	H	H	T	T	H	T	H	T	T	H	H	T	T	H	H	T	T	H	H	T	H	H	H	H	H	T	T	H	H	H	T	H	T	T
H	T	H	H	H	T	T	T	H	H	H	T	T	H	H	T	H	T	T	H	H	T	T	H	T	H	T	T	T	H	H	H	T	T	H	H	T	T	T
H	H	T	H	H	T	T	H	H	T	H	T	H	T	H	H	H	T	T	H	H	H	T	T	H	T	T	H	H	H	H	T	T	H	H	T	T	H	H
T	T	H	T	H	T	H	T	T	T	T	H	T	H	T	H	T	T	H	H	H	T	T	T	H	H	H	H	T	H	H	H	T	H	T	H	T	H	T
T	T	T	T	T	H	T	T	T	H	H	T	T	H	T	T	T	T	T	H	T	H	H	H	H	T	T	H	H	H	H	H	T	T	T	H	T	H	H
H	T	H	H	H	T	H	H	H	H	H	H	T	T	T	T	H	T	T	H	T	T	T	H	H	H	T	H	H	H	T	H	H	H	T	H	T	H	H
H	T	H	H	T	T	H	T	H	H	T	H	H	H	H	H	T	H	T	H	H	T	H	T	H	T	T	H	T	T	H	T	T	H	T	H	T	H	T

In[34]:= **HTrun = {8, 4, 4, 14, 3, 6, 4, 5, 2, 5, 5, 3, 5, 4, 2, 5, 3, 2, 13, 4, 4, 3, 5, 4, 2, 3, 4, 7, 2, 2}**

Q9. The idea is we flip a coin until we get HH and record the # of flips. The question wants you to do this 20 times so you will end up with 20 numbers, each number representing the # of flips to get HH in the corresponding trial. The results you get when you flip a coin is random, so the professor has done the coin flips for you in order to get uniform results (same answers for everyone).

The way to interpret the big matrix of coin flips is:

- each row is a trial (there are 20)
- in each row, you count the number of flips in order to get HH and then discard the remaining flips

There are 20 numbers below the matrix which represent the # of flips to get HH for each trial. The question labels these numbers with X. So the sample mean of X ($\bar{X}$) is just the sample mean of these 20 numbers, and likewise for the sample SD of X (s_x).

$$\bar{X} = \frac{\text{sum of 20 \#'s}}{20} = \frac{96}{20} = 4.8$$

$$s_X = \sqrt{\frac{n}{n-1}}\sqrt{\overline{X^2} - (\bar{X})^2} \qquad \text{where } n = \text{sample size} = 20$$

$$= \sqrt{\frac{20}{19}}\sqrt{37.5 - 4.8^2} = 3.901 \qquad \overline{X^2} = \text{mean of } X^2$$

The easier way is to just plug these #'s into your calculator.

- $\boxed{\text{STAT}}$ → 1: edit
- insert X values into table L_1
- $\boxed{\text{STAT}}$ → CALC → 1: 1-Var Stats → $\boxed{\text{2nd}}$ $\boxed{1}$

$\bar{X} = \bar{x}$ in calc. $= 4.8$

$s_x = S_x$ in calc. $= 3.901$

Q10. Do the same thing as Q9 ~~xxx~~ using the matrix provided.

$\bar{y} = 4.567$

$s_y = 2.849$

Q11. $\bar{X}$ and $\bar{y}$ look about the same. If μ_x and μ_y differed, then μ_x would be bigger since $\bar{X}$ is greater than $\bar{y}$.

Q12. Find a z-based 95% CI for $\mu_x - \mu_y$.

From Q9 and Q10:

$\bar{X} = 4.8$ $\qquad$ $\bar{y} = 4.567$

$s_x = 3.901$ $\qquad$ $s_y = 2.849$

$n_x = 20$ $\qquad$ $n_y = 30$

$$(\bar{X} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

$$4.8 - 4.567 \pm 1.96\sqrt{\frac{3.901^2}{20} + \frac{2.849^2}{30}}$$

$$0.233 \pm 1.991$$

Q13. Find a z-based 95% CI for $p_x - p_y$.

$\hat{p}_x = \frac{20}{30} = \frac{2}{3}$ $\qquad$ $\hat{p}_y = \frac{32}{40} = \frac{4}{5}$

$$(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$$

$$\left(\frac{2}{3} - \frac{4}{5}\right) \pm 1.96\sqrt{\frac{\frac{2}{3}(\frac{1}{3})}{30} + \frac{\frac{4}{5}(\frac{1}{5})}{40}}$$

$$-0.133 \pm 0.209$$

* Note: When we calculate a CI for $\mu_x - \mu_y$ or $p_x - p_y$, we assume that the 2 populations are independent. This is different from the paired data we had when calculating the improved CI for a population mean.

* Note: When we use $t_{\alpha/2}$ for our t-based CI, we assume that the population is roughly normal.

* Importante: If you missed last recitation, get the notes from someone in class. We covered some stuff critical for exam 3.