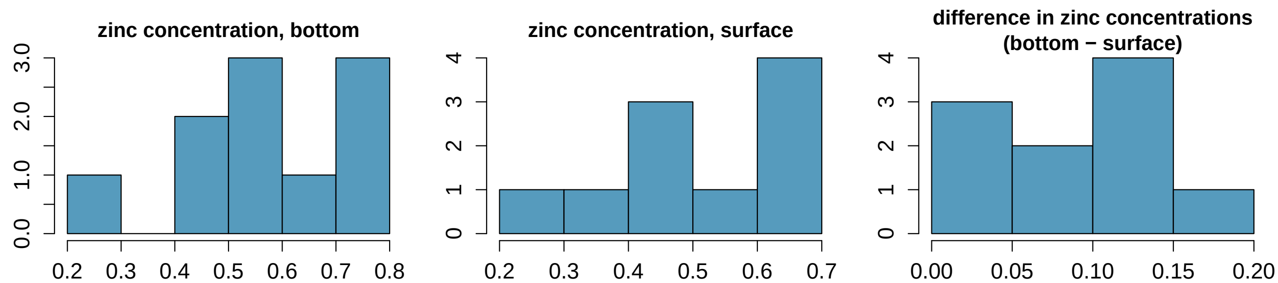
Sergio I Garcia-Rios
Cornell University

Application exercise 4.1: Zinc levels

Team names: ______________________________

Write your responses in the spaces provided below. WRITE LEGIBLY and SHOW ALL WORK!
Concise and coherent are best!

1. Trace metals in drinking water affect the flavor and an unusually high concentration can pose a health hazard. Ten pairs of data were taken measuring zinc concentration in bottom water and surface water at 10 randomly sampled locations. The distributions are shown below. We want to evaluate whether the true average concentration in the bottom water *exceeds* that of surface water? Note that water samples collected at the same location, on the surface and in the bottom, cannot be assumed to be independent of each other. The differences are calculated as $bottom - surface$.



	$\bar{x}$	s	n
bottom	0.5649	0.1468	10
surface	0.4845	0.1312	10
diff	0.0804	0.0523	10

(a) Define the parameter of interest and the point estimate, and state the value of the point estimate.

Parameter of interest: Average difference between the zinc concentrations between bottom and surface water.
Sample statistic: Average difference between the the zinc concentrations between bottom and surface water in the observed data.

(b) Conduct a hypothesis test answering the research question. Don't forget to check conditions first. Use $\alpha = 0.05$. Make sure to frame your conclusion in context of the data and the research question.

$H_0 : \mu_{diff} = 0$
$H_A : \mu_{diff} > 0$
Conditions:

i. *Independence: If the locations are sampled randomly one location can be assumed to be independent of another.*

ii. *Skew: The distribution of differences is not extremely skewed.*

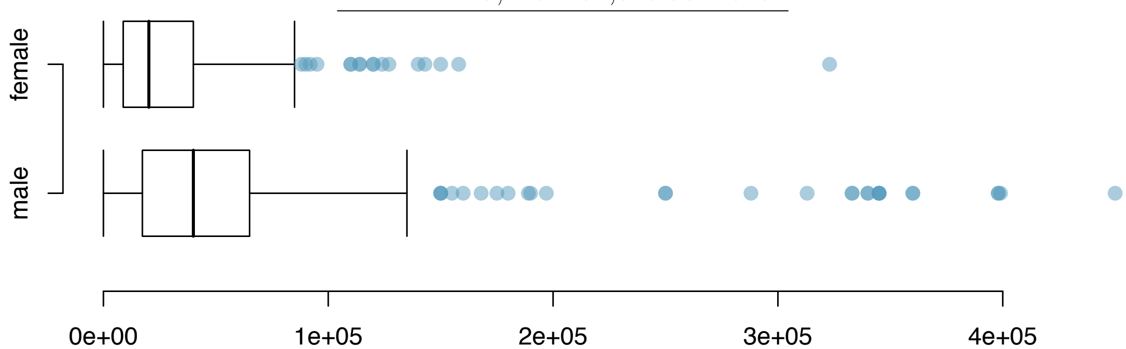
Hence we can assume that the sampling distribution of the average difference in zinc concentrations between bottom and surface water is nearly normal. $T_{10-1} = \frac{0.0804-0}{\frac{0.0523}{\sqrt{10}}} = 4.86$
p-value $< 0.005 \rightarrow$ *Reject* H_0*, the data provide convincing evidence of a difference between in zinc concentrations between bottom and surface water.*

(c) Calculate a confidence interval for the parameter of interest at the confidence level equivalent to the previous hypothesis test. Make sure to interpret the interval in context of the research question.

CL: $1 - 0.05 * 2 = 0.90$ $0.0804 \pm 1.83\frac{0.0523}{\sqrt{10}} = 0.0804 \pm 0.03 = (0.0504, 0.1103)$ *We are 90% confident that the zinc concentration in bottom water is 0.0504 to 0.1103 higher than surface water.*

2. Since 2005, the American Community Survey polls ~3.5 million households yearly. The following summarizes distribution of salaries of males and females from a random sample of individuals who responded to the 2012 ACS:

	$\bar{x}$	s	n
male	55,890	68,767.88	470
female	29,240	32,025.98	373



We want to evaluate whether salaries of men and women are different, on average.

(a) Define the parameter of interest and the point estimate, and calculate the point estimate.

Parameter: Difference in average salaries of all US males and females.
Statistic: Difference in average salaries of sampled males and females.

(b) Conduct a hypothesis test answering the research question. Don't forget to check conditions first. Use $\alpha = 0.10$. Make sure to frame your conclusion in context of the data and the research question.

$H_0 : \mu_m - \mu_f = 0$
$H_A : \mu_m - \mu_f \neq 0$
Conditions:

i. *Independence: The individuals are randomly sampled and are less than 10% of their respective populations.*

ii. *Skew: The distributions are likely right skewed, but we have large sample sizes.*

Hence we can assume that the sampling distribution of the average difference in salaries of males and females are independent of each other.
$T_{371} = \frac{(55{,}890-29{,}240)-0}{\sqrt{\frac{68{,}767.88^2}{470}+\frac{32{,}025.98^2}{373}}} = \frac{26{,}650}{3{,}579.318} = 7.45$
p-value $< 0.01 \rightarrow$ *Reject* H_0*, the data provide convincing evidence of a difference in salaries of males and females.*

(c) Calculate a confidence interval for the parameter of interest at the confidence level equivalent to the previous hypothesis test. Make sure to interpret the interval in context of the research question.

$(55,890 - 29,240) \pm 1.97 \times 3,579.318 = (19,598.74, 33,701.26)$ *We are 95% confident that men on average make 19,598.74 to 33,701.2 more than women.*