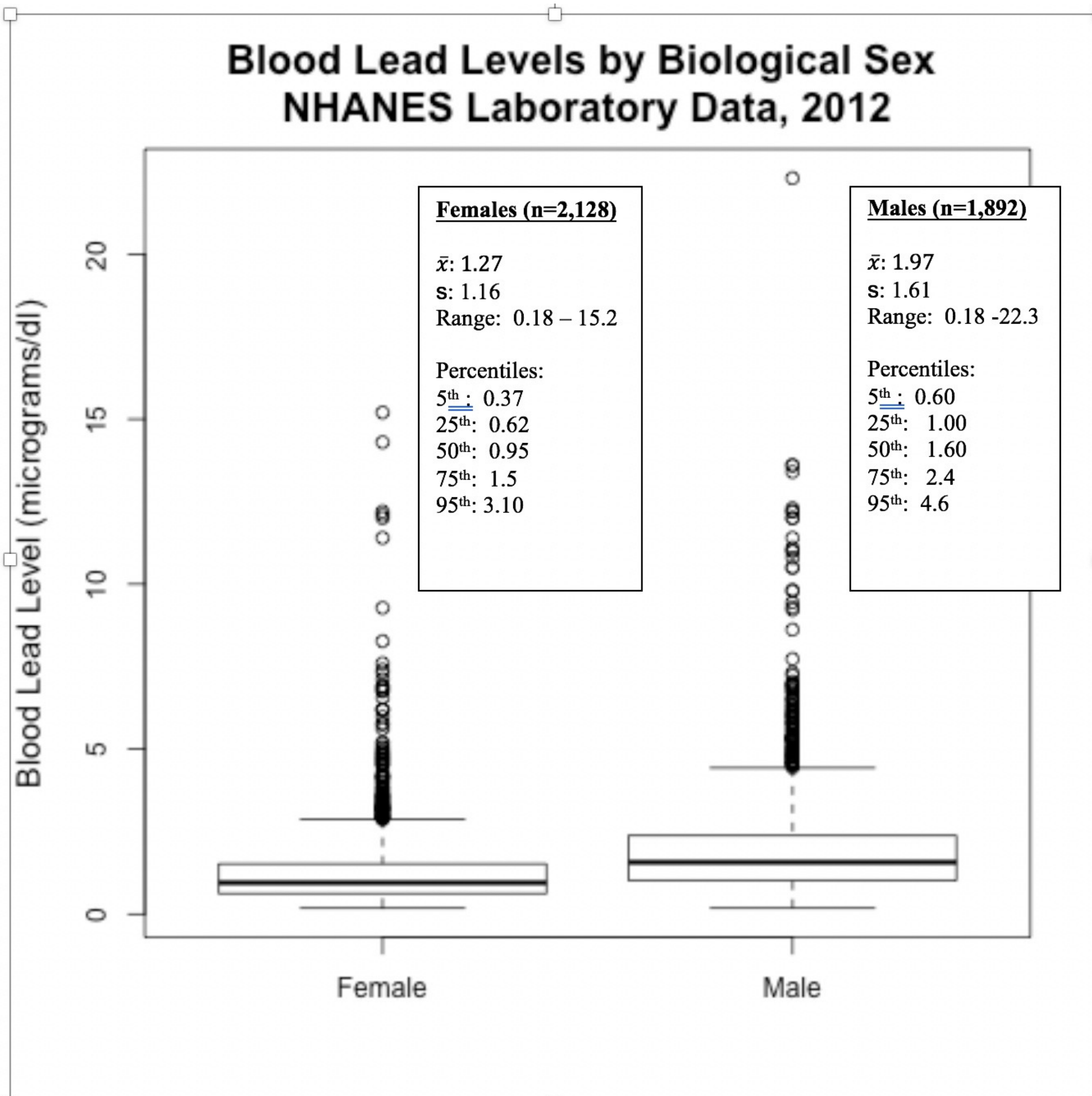
Data from the year 2012 National Health and Nutrition Examination Survey (NHANES) survey includes laboratory measurements on a random sample of more than four-thousand 18-65 year old persons in the United States. The following boxplots display the distribution of blood lead levels(mg/dL) for this sample, separately for males and females. The accompanying tables display summary statistics of the sample values.



1. Suppose you were able to take a random sample of 100 females from the same population of females as the sample of 2,128. How would the sample standard deviation estimate for these 100 sample blood lead level measurements (s) compare, in value, to the sample standard deviation of the original 2,128 blood lead level measurements (1.16 microgram/dL)?

   a. $s_{100} > 1.16$
   b. $s_{100} < 1.16$
   c. s_{100} will be exactly equal to 1.16
   d. s_{100} should be similar in value to 1.16, but there is no way to predict exactly how the two estimates will compare numerically

2. Which of the following best describes the shape of the lead level distributions in both sex groups?

   a. Left-skewed
   b. Symmetric
   c. Right-skewed
   d. Approximately normal

3. What is the mean difference in blood lead levels for females compared to males?

   a. 0.7 micrograms/dL
   b. -0.7 micrograms/dL
   c. 0.45 micrograms/dL
   d. -0.45 micrograms/dL

This sample mean difference is $\bar{x}_{females} - \bar{x}_{males} = 1.27 - 1.97 = -0.7$ micrograms/dL.

4. Suppose, based on these results, researchers decide to make the cutoff for high blood lead levels at 3.10 micrograms/dL. A binary variable is created such that a value of 1 indicates that an individual's blood lead level is greater than 3.10 micrograms/dL, and a value of 0 indicates that an individual's HDL blood lead level is less than or equal to 3.10 micrograms/dL. What percentage of the females would have a value of 1 for this binary indicator? (you may assume there are no repeated values in these data)

   a. 2.5%
   b. 16%
   c. 95%
   d. 5%

The 95th percentile of blood lead levels for females is 3.10 micrograms/dL. As such, 5% of the sample of females have blood lead levels greater than this value

5. The relative risk of having high blood lead levels (greater than 3.10 micrograms/dL) for *males compared to females* is 3.0. What percentage of males have high blood lead levels?

   a. 15%
   b. 10 %
   c. 1.5 %
   d. - 0.7 %
   e. 95%

The relative risk given for males compare to females $= \frac{\hat{p}_{males}}{\hat{p}_{females}} = 3$, where $\hat{p}_{males}$ is the sample proportion of males with blood lead levels greater than 3.10 micrograms/dL. However, from the previous question, we know $\hat{p}_{females} = 0.05$. (5%). So $\hat{p}_{males} = \hat{p}_{females} \times 3 = 0.05 \times 3 = 0.15$, or 15%.

An article from the *American Journal of Public Health* reports the results from a randomized study designed to evaluate the efficacy of an intervention targeted to Hispanic/Latino men who identify as gay, bi-sexual or other men who have sex with men (MSM). A representative sample of 254 such men was randomized to be in either the intervention group (n=152) or the control group (n=152). The primary outcome under study getting tested for HIV within the six-months following group assignment (randomization) among those who had been sexually active in this same six-month follow-up period.

At six months of follow-up, 141 subjects in the intervention group reported having had sex (with men and/or women) since randomization. Of these 141 men, 114 had been tested for HIV since being randomized. At six months of follow-up, 147 subjects in the control group reported having had sex (with men and/or women) since randomization. Of these 147 men, 40 had been tested for HIV since being randomized.

6. What is the (approximate) estimated **relative risk** of getting tested for HIV for subjects in the *Intervention group* compared to subjects in the *Control group*?

   $\widehat{RR} = \frac{\hat{p}_{Intervention}}{\hat{p}_{control}} = \frac{(114/141)}{(40/147)} = 2.97 \approx 3$. Answers in the range of 2.9 to 3.1 receive full credit.

A 2016 article in *JAMA* reports the results of a study of treatment outcomes for children with mild gastroenteritis who were given oral rehydration therapy. Enrolled children were randomized to received either rehydration with diluted apple juice (DAJ), or an electrolyte maintenance solution (EMS). As per the study authors:

> "The primary outcome was a composite of treatment failure defined by any of the following occurring within 7 days of enrollment: intravenous rehydration, hospitalization, subsequent unscheduled physician encounter, protracted symptoms, crossover, and 3%or more weight loss or significant dehydration at in-person follow-up. Secondary outcomes included intravenous rehydration, hospitalization, and frequency of diarrhea and vomiting."

Of the 323 children randomized to DAJ, 54 experienced treatment failure. (17 %). Of he 324 children randomized to EMS, 81 experienced treatment failure. (25 %)

7. Estimate the risk difference (difference in proportions) of treatment failure for children in the DAJ group compared to children in the EMS group. (DAJ-EMS). Please report as a decimal, not a percentage.

$$\widehat{RD} = \hat{p}_{DAJ} - \hat{p}_{EMS} = 0.17 - 0.25 - \; -0.08.$$

8. Estimate the relative risk (risk ratio) of treatment failure for children receiving DAJ compared to children receiving EMS.

$$\widehat{RR} = \frac{\hat{p}_{DAJ}}{\hat{p}_{EMS}} = \frac{017}{0.25} = 0.68.$$

An article from the *American Journal of Public Health* reports the results from a randomized study designed to evaluate the efficacy of an intervention targeted to Hispanic/Latino men who identify as gay, bi-sexual or other men who have sex with men (MSM).

A representative sample of 254 such men was randomized to be in either the intervention group (n=152) or the control group (n=152). The primary outcome under study getting tested for HIV within the six-months following group assignment (randomization) among those who had been sexually active in this same six-month follow-up period.

men and/or women) since randomization. Of these 141 men, 114 had been tested for HIV since being randomized. At six months of follow-up, 147 subjects in the control group reported having had sex (with men and/or women) since randomization. Of these 147 men, 40 had been tested for HIV since being randomized.

9. The difference in proportions of men being tested in the intervention group compared to the control group is 0.54 (54%). Suppose this intervention were used in a community with 1,000 Hispanic/Latino men who identify as gay, bi-sexual or other men who have sex with men (MSM). What would be the expected effect on HIV testing outcomes?

   a. There would be an estimated 460 fewer men getting tested for HIV (in the six months following the intervention) than if the intervention was not given.
   b. There would be an estimated 540 fewer men getting tested for HIV (in the six months following the intervention) than if the intervention was not given.
   c. There would be an estimated 460 more men getting tested for HIV (in the six months following the intervention) than if the intervention was not given.
   d. There would be an estimated 540 more men getting tested for HIV (in the six months following the intervention) than if the intervention was not given.

The estimated risk difference (difference in proportions) of persons getting tested in the intervention group compared to the control group is $\widehat{RD} = \hat{p}_{int} - \hat{p}_{control} = 0.54$ or 54% - i.e., on the absolute scale 54% more men got tested in the intervention group. This estimated can be applied to a community of a specific size to get an estimate of the (in this case) increased number of outcomes (tests) that would be expected if all were given the intervention: 0.54*1,000 = 540.

10. Consider studies designed to compare the occurrence of a binary outcome between two populations: population A and population B. In general, which of the following statements best describes the relationship between the relative risk estimate ($R\hat{R}$)and the odds ratio estimate ($O\hat{R}$), both based on the same two samples from populations A and B?

   a. $R\hat{R}$ and $O\hat{R}$ will always be exactly the same in value.
   b. If $R\hat{R}>1$ then $O\hat{R}$ will be less than 1.
   c. $R\hat{R}$ and $O\hat{R}$ may differ in value, but will show the same direction of association.
   d. $R\hat{R} = \dfrac{O\hat{R}}{\sqrt{n_1 + n_2}}$, where n_1 and n_2 are the sizes of the samples from population A and population B, respectively