* Researchers are studying relative concentration of blood lead for factory workers. They took the natural log of ratio of blood lead concentration for 8 workers + 8 control subjects. Measurements resulted in a mean log concentration = 6 (log parts per volume) for workers + 4 for controls. Sample variance in workers = 3 + = 5 in the controls. *Assuming equal variances*, create a 95% CI for the difference in population means of log blood lead concentration between factory workers + controls (Factory works - Controls)
* **[0.04, 3.96] (log parts per volume)**
* 

**n\_w <- 8**

**n\_ctrl<- 8**

**mean\_w <- 6**

**mean\_ctrl <- 4**

**var\_w <- 3**

**var\_ctrl <- 5**

**df <- sum(n\_w,n\_ctrl)-2**

**std\_pooled\_var\_estimator <- ((n\_w - 1)\*3 + (n\_ctrl - 1)\*5)/df**

**## grab quantiles for 95% CI**

**qtiles <- qt(p=.975,df=df)**

**# compute CI bounds**

**(lo <- (mean\_w - mean\_ctrl) - qtiles\*sqrt(std\_pooled\_var\_estimator\*((1/n\_w)+(1/n\_ctrl))))**

**(hi <- (mean\_w - mean\_ctrl) + qtiles\*sqrt(std\_pooled\_var\_estimator\*((1/n\_w)+(1/n\_ctrl))))**

* Suppose the interval was [-0.2, 3.0] log parts per volume. What would this say about factory workers versus controls?
* **Cannot rule out possibility the true means are equal at 95% confidence b/c 0 is in the interval**
* Let S12, S22 and S32 be sample variances from random samples of size n1, n2 + n3 respectively. The populations have means μ 1, μ2 + μ3 respectively, + a common variance δ2. When is  unbiased?
* **For any numbers so that π1 + π2 + π3 = 1**
* 
* Do *NOT* have to be positive for this estimate to be unbiased.
* *However, if there IS a negative, the resulting variance estimate has a non-zero probability of being negative.*
* In an effort to improve efficiency, hospital admins are evaluating a new triage system for the ER. In an validation study of the system, 5 randomly selected out of 10 patients were tracked in a mock ER w/ the old triage system while the remaining ones were tracked under the new triage system. Waiting times were recorded for all 10 patients. Would it be better to use an independent group or paired T confidence interval in this setting?
* **Independent group T interval**
* Suppose you create a 95% CI + later decide to create a 99% CI. What can be said about the new interval with respect to the 95% interval?
* **The new interval will be wider.** 
* You have 3 DP’s {1, 3, 7}. What is the exact bootstrap distribution of the sample median? (Note - specifically - the bootstrap distribution of the median.)
* **Distribution that takes 1 w/ p =7/27, 3 w/ p =13/27 and 7 w/ p =7 / 27**
*  **= empirical distribution**
* **temp <- expand.grid(c(1, 3, 7), c(1, 3, 7), c(1, 3, 7))**
* **table(apply(temp, 1, median))**
* 1 3 7
* 7 13 7
* Suppose we simulated a large amount of random uniform #’s + a large amount of exponential(1) #’s. What would a plot of the quantiles of 1 vs. the other look like? (Uniform on the horizontal axis + the exponential on the vertical axis.)
* **It would look like a plot of the function f(x) = log(1 - x)**
* Let *X* = a uniform random variable then the **α**th quantile = **α**
* For an exponential, this is **−log(1−*α*)**
* Consider any distribution function *F* w/ associated mean 0 + variance 1. Notice that if *X* is a random variable from F, then **Y = μ + Xδ** has mean **μ** and variance **δ2**. What is the distribution function associated with Y? (Call it G.)
* 
* Let F and G be distribution functions. If you do a QQ plot w/ quantiles of F on the horizontal axis + quantiles of G on the vertical axis, what must it look like?
* **The function ** where G-1 = quantile function of G
*  🡺 plotting quantile of G as function of F
* Consider the points in a QQ plot, going from left to right in the plot, can a point ever be lower than the previous point?
* **No 🡪 both the empirical and theoretical quantiles must be non-decreasing.**
* In a new population, a sample of 9 men yielded a sample average brain volume of 1,100cc + SD of 30cc. What is a 95% Student's T CI for the mean brain volume in this new population?
* **About 1,077 to 1,123 cc**

**x\_bar <- 1100**

**s <- 30**

**sample\_var <- s\*\*2**

**n <- 9**

**stdErr <- s/sqrt(n)**

**alpha <- .05**

**## grab quantiles**

**qtiles <- qt(p =1-alpha/2, df = n-1)**

**## rev() = reverse elements due to order of returned qtiles**

**(lo <- x\_bar - qtiles\*stdErr)**

**(hi <- x\_bar + qtiles\*stdErr)**

* A diet pill is given to 9 subjects over 6 weeks. The average difference in weight (follow up - baseline) is -2 pounds. What would the SD have to be for a 95% T confidence interval to lie entirely below 0?
* **Around 2.6 pounds or less**

**x\_bar <- -2**

**n <- 9**

**alpha <- .05**

**## grab quantiles**

**qtiles <- qt(p =1-alpha/2, df = n-1)**

**(x\_bar/qtiles)\*sqrt(n) = -2.6**

* The interval would up being [-3.5, -0.5] pounds. What can be said about the population mean weight loss at 95% confidence?
* **There is support at 95% confidence of mean weight loss.**
* In an effort to improve efficiency, hospital admins are evaluating a new triage system for their ER. In an validation study of the system, 5 patients were tracked in a mock ER under both the new and old triage system. Their waiting times were recorded. Which t CI should be used in this setting?
* **A paired interval**
* To further test the system, admins selected 20 nights + randomly assigned the new triage system to be used on 10 nights + the standard system on the remaining 10 nights. They calculated nightly median waiting time (**MWT**) to see a physician. Average MWT for the new system = 3 hours w/ variance = 0.60 while average MWT for the old system = 5 hours w/ variance of 0.68. Give a 95% CI estimate for the differences of mean MWT associated w/ the new system. *Assume constant variance*
* **A mean decrease of between 2.75 and 1.25 hours.**
* Suppose that you create a 95% t-CI. You then create a 90% CI using the same data. What can be said about the 90% CI with respect to the 95% CI?
* **The interval will be narrower.**
* Let distribution 1 be N(μ1,σ21) and distribution 2 be N(μ2,σ22). Let x1,α and x2,α be the αth quantile from the two distributions, respectively. How are the two mathematically related?
* **They are related as a line **
* Consider data points x1, …, xn. Imagine a PMF so that a random variable X from this distribution has == the so-called bootstrap distribution. What is the mean of this distribution?
* **The sample mean of the data**
* Suppose we were to simulate a large # of standard normal random variables + a large # of exponential random variables. What would a plot of the exponential quantiles (horizontal axis) versus the standard normal quantiles (vertical axis) look like? Let Φ be the standard normal distribution function.
* **It will look like the function **
* Let F(x) be a distribution function. Notice G(x)=F(a2x+b) is also a distribution function for a≠0. If you were to take large samples from F and G, what must the QQ plot look like without knowing the specific values of a and b?
* **A line.**
* Let your data be the 2 points {1,3}. What is the bootstrap distribution of the sample mean?
* **The distribution puts probabilities 1/4, 1/2 and 1/4 on the numbers 1, 2 and 3, respectively**
* **temp <- expand.grid(c(1, 3), c(1, 3))**
* **table(apply(temp, 1, mean))**
* **1 2 3**
* **1 2 1**