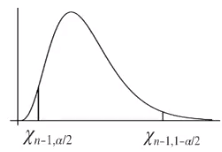
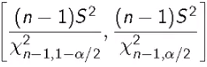
**Confidence Intervals and CI for Normal Variance**

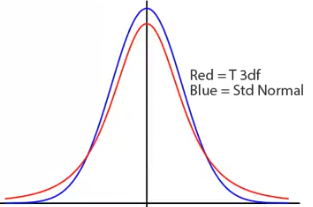
* This is based on the assumption our distribution = Gaussian
* Can use CLT to create CI’s, + to make better ones for smaller samples using **Gosset’s/Student’s t-distribution** (treating the data as if continuous)
* **General Procedure for CI** (creating a probability statement + manipulating it to generate an interval)
* Create a **pivot/**statistic whose distribution does NOT depend on the parameter of interest
* Ex: Use CLT+ take a sample mean, subtract off population mean you're interested in, + divide by standard error, that statistic *clearly* depends on the parameter of interest.
* BUT the *distribution* of that statistic, at least in the limit, *doesn't* depend on the parameter you're interested in the sample mean.
* Solve for probability that the pivot lies between bounds of the parameter of interest
* Must also know **Chi-squared distribution**
* Suppose **S2 = sample variance** from collection of **IID N(µ,σ2) data**, then \*\*\*
* i.e. n - 1, times sample variance, divided by variance, gives a **Chi-squared random variable, χ2n-1, w/ n-1 degrees of freedom**
* i.e the **normalized sample variance** follows a **Chi-squared distribution w/ n-1 dF**
* Chi-squared distribution = skewed w/ support on 0 to Inf, its mean = its dF, + its variance = 2\*dF:
* **E[χ2dF] = dF Var(χ2dF) = 2dF**
* Recall: sample variance = unbiased estimator (why we divide by n, not n-1)
* 
* Note that if  is the α quantile of the Chi-squared distribution, then**:**
* 

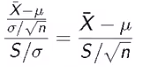
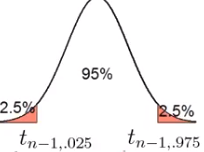
Chi-square density: 

* i.e. Probability our chi-squared random variable  is between the 2 quantiles defined by α must be == 1 – α
* Here,  = our pivot
* Then solve for parameter of interest: 
* i.e. we have a 1 – α probability that the random interval given contains our variance
* So that  == a **100(1-α)% CI for σ2**
* α could be .05, so we wind up w/ a 95% CI for our parameter of interest, σ2
* **In a CI, the \*\*\*interval is random, and the parameter is fixed**\*\*\*
* When we collect data + form the CI, it either contains the parameter or not (no probability in this statement, its either 1 or 0)
* “Intro Stat” interpretation of a CI = a procedure that if you were to repeatedly do the experiment + form CI’s, 95% of the CI’s would contain the parameter of interest
* Much weaker interpretation = you get 2 numbers that = an interval estimate of the parameter to estimate, but said interval estimate incorporates uncertainty.
* Notes about the interval from above (**100(1-α)% CI for σ2**)
* \*\*\*This CI relies *heavily* onassumed normality + is not too robust to departures from normality\*\*\*
* We *could* get the CI in other ways, if not normal, such as **bootstrapping**
* Square-rooting the endpoints = CI for **σ** (SD) = 
* Turns out that **(n-1)S2 ~ Gamma{(n-1)/2, 2σ2}** == “follows a gamma distribution w/ **shape** = (n-1)/2 and **scale** = 2σ2”
* Therefore, this can be used to plot the likelihood function for σ2
* Bit difficult since underlying data is Gaussian w/ 2 parameters µ, σ2, so our likelihood is a **bivariate function** (µ on 1 axis, σ2 on the other, likelihood as the height)
* Could create a **marginal likelihood** for σ2
* \*\*\*If we *don’t* divide by σ2, then (n-1)S2 still contains its units (dividing by σ2 makes it unitless, which is a requirement of the Chi-squared distribution, which is unitless\*\*\*)
* If we *don’t* divide by σ2, we end up w/ a Gamma distribution which is indexed by 2 parameters (**shape, scale**) as seen above
* So, we have data + a single number, (n-1)S2, + if we’re willing to assume the DP’s that make up this number are Gaussian, we can take the Gamma density, plugin the data, view it as a function of the parameters, + plot a likelihood function
* Ex: Study of 513 organo-lead manufacturing workers reported an average TBV of 1150.315 cm3 w/ **σ =** 105.997. Assuming normality of the underlying measurements, calculate a CI for population variation in TBV
* **## Create CI for variance**
* **x\_bar <- 1150.315**
* **s <- 105.997**
* **sample\_var <- s\*\*2**
* **n <- 513**
* **alpha <- .05**
* **## grab quantiles**
* **qtiles <- qchisq(p = c(alpha/2,1-alpha/2), df = n-1)**
* **## rev() = reverse elements due to order of returned qtiles**
* **(pop\_var\_CI <- rev((n-1)\*sample\_var/qtiles))**
* **[1] 9976.751 12749.451**
* **# interval for sd**
* **(pop\_sd\_ci <- sqrt(pop\_var\_CI))**
* **[1] 99.88369 112.91347**
* This interval (100,113) is created in a way that if assumptions of the interval are correct (data are IID Normal w/ fixed mean **µ** and fixed variance **σ2**), then if we repeat this procedure over and over and over, 95% of intervals obtained would be intervals that contain the true parameter we’re trying to estimate (variance or SD)
* Can plot likelihood:

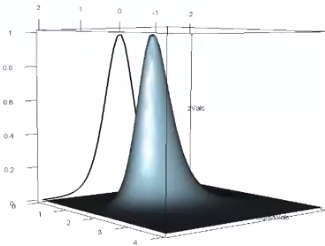
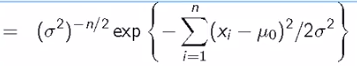
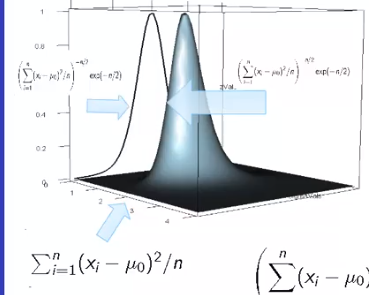
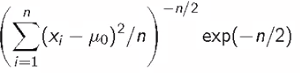
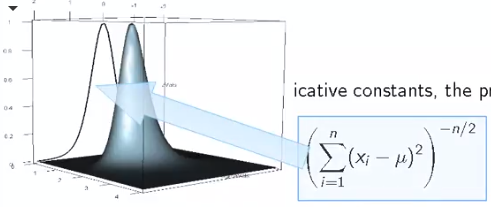
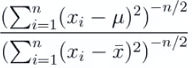
**Student's t Distribution and CI for Normal Means**

* Invented by Williams Gosset (published under pseudonym “Student”
* Has thicker tails than Gaussian, is indexed by **degrees of freedom** + becomes more like the Standard Noemal as dF grows

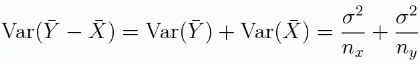
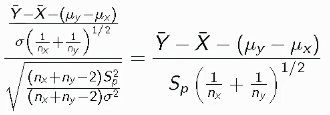
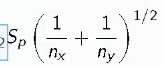
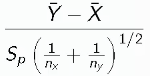
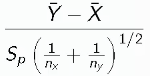


* Obtained via  where Z + χ2 = independent standard Normals + independent Chi-squared distributions, respectively
* Suppose (X1, …, Xn) are IID N(**µ**,**σ2**) then:
* 1) t-statistic =  is standard normal
* b/c linear combinations of normal random variables are themselve normal (i.e. Xbar = normal)
* b/c they’re IID, we know exactly what **SD of Xbar** is (**sigma / sqrt(n)**) + we know its mean = **µ**
* So, when we shift + scale our *nonstandard* normal by **µ** + divide by its SD, we get a standard normal
* 2)  = square root of a Chi-squared divided by its dF
* Therefore, take the 2 values above 🡺  (take a Chi-squared + divide by its dF) will follow Gosset’s t-distribution w/ n-1 dF
* Haven’t *exactly* shown that Xbar and S are independent (they’re from the same data) 🡺 assume for now
* Previously, in constructing CI’s, **Xbar -** **µ** divided by σ/n is a nice pivotal statistic that’s useful for generating CI’s (+ also for doing hypothesis tests.
* All we've done here is replaced σ w/ S 🡺 \*\*\*can take the *unknown population variance* + replace it w/ the *known sample variance* + weget a statistic whose distribution we know*\*\*\**
* This statistic, , also limits to a Standard Normal as X goes to infinity
* Ex: Create CI for the mean
* **t statistic** is a **pivot**, so (*under the assumption of normality of the data*), does NOT depend on parameter of interest (**µ**) + therefore can be used to create a CI for parameter of interest (**µ**)
* \*\*\*Let  =  quantile of the t-distribution w/ **dF** degrees of freedom\*\*\*
* Then, 1 – α =  = probability our t-statistic/random variable lies w/in our CI is true
* Rearrange terms to get CI for **µ** = 
* So, our \*\*\*interval =  == **estimate +/- quantile \* std. error**\*\*\*
* Notes about t-interval
* t-interval *technically* assumes data are IID Normal, though it is robust to this assumption
* Works well whenever distribution of the data is *roughly*  symmetric + mound-shaped
* Paired observations (ex: patients before + after treatment) are often analyzed using a t-interval by taking differences (which tend to be much more Gaussian)
* For large dF, t-quantiles becomes the *same as Standard Normal quantiles*
* Therefore, this interval converges to the same interval as from the CLT
* For skewed distributions, t-interval assumptions are violated
* Also, for skewed distributions, doesn’t make sense to center interval @ the mean
* In this case, consider taking **logs** or using a different summary stat (i.e. median)
* For *highly discrete data* (like binary), other intervals are available
* Can use t-distribution to create a likelihood for a single parameter (which itself is function of 2 parameters, **µ and σ2**)
* If X is N(µ,σ2) + χ2 is a Chi-squared random variable w/ dF degrees of freedom, then  = **non-central t random variable w/ non-centrality parameter = µ/σ**
* µ/σ = the mean in SD units 🡺 unit-free quantity called the **effect size**
* Useful for creating **likelihood effect size**
*  == did not subtract off µ, so X/σ still has a mean (= µ/σ), so we have NOT taken a Standard Normal + divided by the square root of an indepentn Chi-squared divided by its dF, but instead took a NON-Standard Normal + divided by the square root of an indepentn Chi-squared divided by its dF
* So, it *can’t* work out to be a t-random variable b/c we haven’t satisfited the definision of a t-random vairble, so that’s why it’s a **non-central t random variable**
* But when µ = 0, we DO have a normal t random variable
* Non-central t random variable has a dF, but also has a 2nd parameter, the non-centrality parameter µ/σ
* For context, Xbar is Normal(µ, σ2/n) and (n-1)S2/σ2 = chi-squared w/ n-1 dF
* Then,  = a non-central t w/ non-centrality parameter = 
* This can be used to create a 1D likelihood for **effect size =** µ/σ w/out any further tricks

**Profile Likelihoods**

* **Profiling** = preferred method for univariate likelihoods from multivariate likelihoods
* We’re going to look at the bivariant Normal distribution which has 2 parameters µ, σ2, + we’re going to obtain a likelihood for µ alone (can also obtain a likelihood for σ2 alone)
* **Profile Likelihood (PL)** gets name b/c results = shadow we’d get if we shined a light on the 2D likelihood for µ and σ2
* PL is a bivariant surface w/ µ and σ2 on the axis + height = likelihood
* Ex: for shining light along σ axis
* Shadow = likelihood placed on the plane defined by the µ direction
* In other words, we want to get the function obtained onto the wall where the shadow occurs
* PL for parameter value µ0 = obtained by maximizing **joint likelihood** for σ w/ µ fixed @ µ0
* *Finding value of the curve in the shadow @ µ0*
* Light will go through all values above the likelihood + get stopped anywhere on the likelihood + up until the maximum value.
* This process is then repeated for many values of µ0
* Ex: **joint likelihood w/** µ fixed @ µ0  == the Gaussian density , and since we have independent data we take a product out from == 
* Collect all terms == 
* Having µ0 fixed, MLE for σ2 🡺 **log** the likelihood + take derivatives + solve for σ2 🡺 (generalization of the variance)
* Remember to fix σ2 as the parameter we’re deriving w/ respect to, not σ
* So, if we fix µ @ a particular value, our MLE for σ2 = sample variance but instead of subtracting deviations from the sample mean, we’re subtracting deviations from our fixed value for σ2
* == peak of our likelihood == the point where the light switches from *not* going through the likelihood to the point right above it where the light passes over the likelihood, which is the point that gets shadowed onto the wall @ µ0
* 
* Plugging this peak back into likelihood again, we get 
* This e(-n/2) is irrelevant since it doesn’t involve µ0
* All the above is done for 1 value of µ0, so if we do it for every µ0 then we’d get a function that = our profile likelihood 🡺 
* **profile likelihood** = a summation of x minus µ squared raised to -n/2 over all n
* **this function is clearly maximized at** 
* In general, a nice property of the profile likelihood is the **maximizer of the profile likelihood/maximum profile likelihood estimand is also the MLE for the parameters**
* In this case  is the same as the MLE for parameter µ for the complete likelihood
* If we wanted to divide this by its peak value, divide it by the same thing w/ x\_bar plugged in for µ to normalize the function to tap out @ a value of 1 🡺 
* We treat profile likelihood as if it was a Standard Univariate Likelihood forµ (higher values = better supported, peak = where MLE occurs, horizontal lines to get likelihood-based intervals for µ)
* **NOTE:** The specific technique of the t-CI is a very robust interval, as long as the data look roughly mount-shaped
* the t-CI + the Standard Normal CI look the same, except w/ the t-quantile replaced by a standard normal quantile
* t-CI limits to the standard normal CI, so just always do a t-CI (don’t worry about which sample size to switch between) 🡺 Just never do a standard normal CI + then you don't even have to worry about it
* ***If your sample size is big enough, the t-quantile looks like a normal quantile anyway***

**t-Confidence Intervals**

* **Independent groups t-CI’s**
* Want to compare 2 mean BP between 2 groups in a randomized trial (received Tx vs. placebo)
* Cannot use **paired** t-test b/c groups = independent + may have different sample sizes
* Let both **X1,…,Xnx+ Y1,…Yny** be **IID N(μx, δ2) + N(μy, δ2)** respectively, **Xbar, Ybar, Sx, Sy** be the means **+ SD’s**
* Using the fact that **\*\*\*linear combinations of normals are again normal\*\*\*,** we know **Ybar-Xbar** (*what we want to estimate w/ some uncertainty)*isalso normal w/ **mean = μx + μy** and **variance = δ2(1/nx+ 1/ny)**
* 
* **The \*\*\*Pooled variance estimator\*\*\* =** S2p = is used as a good estimator (*an MLE*) of **δ2**
* Note:  == i.e. complements
* So, S2p = a weighted average of the 2 group variances (for group X and group Y)
* If we have the same sample sizes (nx= ny), then ends up being ½, in which case the pooled variance estimate ends up being the arithmetic averages of the 2 variances
* 
* But if X contains much more data, nx – 1 >>> ny – 1, so the numerator in π above gets much larger, so the weight on S2xover S2y is much larger
* Then, the weighted averages takes whichever group has more measurements + weights the variance from that group more heavily (which is what we want)
* i.e. more data = good estimator places more weight = variance estimated better
* **Pooled variance estimator** = mixture of group variances w/ greater weights on variances from groups w/ larger sample sizes
* If sample sizes are =, Pooled variance estimator = normal average of group variance
* **\*\*\*Pooled variance estimator is unbiased\*\*\***
* Take expected value of the PVE + use the fact that the individual group variance estimators are unbiased 🡺  
* Pooled variance estimator is also independent of Ybar-Xbar since Sx is independent of Xbar and Sy is independent of Ybar
* Recall: **\*\*\*sum of 2 independent Chi-squared random variables is Chi-squared w/ dF = sum of the dF in the summands**
* Therefore, 
* 1st term is chi-squared w/ nx-1 dF, and 2nd = chi-squared w/ ny-1=1 dF 🡺 
* From the fact in bold above + b/c our random variables are independent 
* Would like to create a t-CI, which needs a standard normal divided by the square root of an independent chi-squared which itself is divided by its dF
* Need some function of the PVE to be Chi-squared (?)
* Now construct t 🡺 
*  == standardizing ( = (value – mean) / std. err) == results in a Standard Normal
* *Original data are Gaussian == sample means are Gaussian == different in sample means is Gaussian 🡺 take a Gaussian, subtract mean + divide by SD, we get a Standard Normal*
* Divide Standard Normal it by a chi-squared divided by its dF 🡺  🡺 must result in a t random variable w/ (nx + ny – 2) dF
* Collecting terms + working through the math gives the RHS above == observed difference in means minus population difference in means divided by Std. Err (w/ δ replaced w/ out data estimate of δ, Sp)
* Therefore, this t statistic follows Gosset’s t distribution w/ nx + ny – 2 dF
* Like before, we took our statistic, replaced unknown SD w/ its estimate, + what would be a Gaussian random variable turned into a t random variable in form == (estimator–true value) / SE
* Therefore, a (1-α)\*100 CI for **μx - μy is** 
* **This == \*\*\*estimate (Ybar-Xbar) +/- appropriate quantile from appropriate distribution (t-quantile form the t-distribution =** **) \* by SE** \*\*\*
* This interval assumes constant variance across groups (there are tests to check for this = **F-test**, but these are typically bad)
* If doubt in this, assume different variance per group
* Textbooks typically suggest testing equality of the variances + if equal, do one CI + if unequal, do another CI 🡺 bad strategy
* Instead, look at the data (graphs) + make assessments as to whether or not the variances are equal/unequal + use that to decide.
* If you *must* estimate ratio of the variances in the groups, try bootstrapping (unless sample sizes are very small)
* Safe + conservative thing = just always assume the variances are unequa
* How to get a likelihood for **μy – μx**
* Getting a likelihood for **μy – μx** divided by δ (which is still a single parameter) is very easy.
* The reason is that  (Ybar - Xbar divided by its standard error) follows a **non-centrality distribution**, + the **non-centrality parameter** depends on **μy – μx** over δ + something about the n’s that we know 🡺 **non-centrality parameter** == 
* We can use the statistic  ro create a likelihood for **μy – μx / δ (= a standardized measure of the change (difference) in group means, relative to the inter-group SD/in SD units)**
* “in SD units/relative to inter-group SD” = useful for calibrating difference in means across studies
* Ex: Constructing t-CI with a PVE
* Comparing Sys BP for 8 oral contraceptive users vs. 21 controls
* Xbaroc = 132.86 mmHg, w/ soc = 15.34 mmHg
* Xbarc = 124.44 mmHg, w/ sc = 18.23 mmHg
* **PVE s2p** **=** 
* **t27,.975 = (qt.975,df=27) = 2.052**
* **Interval =** 
* **Now, check if the interval contains 0, b/c if so, then a reasonable estimate for the difference in BP between the 2 groups is 0 (i..e they’re identical)**
* **This means there’s evidence that there is no difference/that this oral contraceptive use does not appreat to be presenting evidence of an associate dincrease in BP**