Chapter 7

ST 511 - Inferences About Variances

Readings: Chapter 7 (for 7.4 read if interested)

We saw in the 2-sample t-test we may have interest in testing if two population variances are equal (i.e. $\sigma_1^2 = \sigma_2^2$).

To investigate this, we first start by looking at inference for a single population variance.

Inference for σ^2

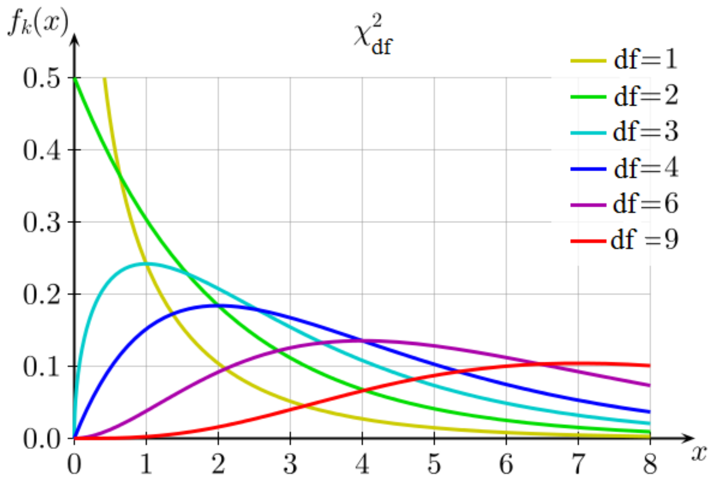
To make inference for σ^2 we need a corresponding statistic...

This is 'unbiased' for σ^2,

To create a CI or HT, we need to know the ____

Theorem: If $Y_i \sim^{iid} N(\mu, \sigma^2)$ (i.e. a RS from a normal parent population) then

Note: Large n will not relax this assumption! We must have the assumed normality here!



Mean = df, Variance = 2(df)

How can we make a $(1-\alpha)100\%$ CI for σ^2?

To get the χ^2_L or χ^2_U values in SAS we can do the following:

```
*Syntax     PROBCHI(x,df)        P(Chi^2_df <x) = returned value
The The PROBCHI function returns the probability that an observation from a chi-square distribution, with
degrees of freedom df is less than or equal to x. This function accepts a noninteger degrees of freedom
parameter df if needed);

*Syntax     QUANTILE(dist, probability, parm-1,...,parm-k)       P(dist<= value returned) = probability
The QUANTILE function computes the probability from various continuous and discrete distributions
'probability' is a numeric constant, variable, or expression that specifies the value of a random variable.
parm-1,...,parm-k are optional shape, location, or scale parameters appropriate for the specific distribution.;


*Find some probabilities and quantile values from a chi-square;
data chisq;
prob1 = probchi(2,2); *Probability chi-sq 2 is less than its mean --- P(Chi^2_2<2);
prob2 = probchi(12.8,4); *P(Chi^2_4<12.8);

quant1 = quantile('chisq',0.95,11); *0.95 quantile from a chi^2_11;
quant2 = quantile('chisq',0.99,15); *0.99 quantile from a chi^2_15;
run;

proc print data=chisq;
title 'Chi-Square values';
run;
```

Chi-Square values

Obs	prob1	prob2	quant1	quant2
1	0.63212	0.98770	19.6751	30.5779

Example: A dairy processing company claims that the variance of the amount of fat in the whole milk processed by the company is 0.25 g^2. You collect a sample of 41 milk containers and find a sample variance of 0.27 g^2. Find a 90% CI for σ^2 = true variance of the amount of fat in the company's whole milk. What do you think of the company's claim? Useful values: $P(\chi^2_{40} > 55.758) = 0.05, P(\chi^2_{40} > 26.509) = 0.95$

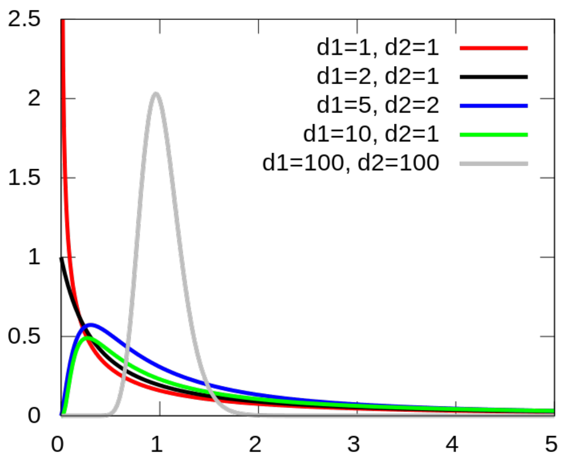
A hypothesis test for $\sigma^2 = \sigma_0^2$ could be done using the test statistic $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$. We won't cover this in class.

Both the CI and the HT rely heavily on the normality assumption. If normality does not hold then the interval and test will not be valid!!! In fact, they perform very poorly (they are not robust to this assumption being violated).

Inference for two variances, σ_1^2 and σ_2^2

Now we are ready to compare two variances (as is needed in the two-sample t-test).

Theorem: If $Y_i \sim^{iid} N(\mu_1, \sigma_1^2)$ $(i = 1, ..., n_1)$ and $X_i \sim^{iid} N(\mu_2, \sigma_2^2)$ $(i = 1, ..., n_2)$ where the Y's and X's are independent then



Notice, when comparing variances we are looking at the ratio $\frac{\sigma_1^2}{\sigma_2^2}$ rather than $\sigma_1^2 - \sigma_2^2$. This is because we know the distribution of the statistic above which involves ratios rather than differences. What value is of interest for this ratio?

How can we make a $(1-\alpha)100\%$ CI for σ^2?

To get the F_L or F_U values in SAS we can do the following:

```
*Syntax    PROBF(x,ndf,ddf)    P(F_df1,df2<x) = returned value
The PROBF function returns the probability that an observation from an F distribution,
with numerator degrees of freedom ndf (our df1) and denominator degrees of freedom ddf (our df2)
is less than or equal to x.

*Find some probabilities and quantile values from an F distribution;
data f;
prob1 = probf(10,4,2); *Probability F_4,2 is less than 10 --- P(F_4,2<10);
prob2 = 1-probf(22,3,18); *P(F_3,18>22);

quant1 = quantile('f',0.95,4,2); *0.95 quantile from an F_4,2;
quant2 = quantile('f',0.99,3,18); *0.99 quantile from an F_3,18;
run;

proc print data=f;
title 'F values';
run;
```

F values

Obs	prob1	prob2	quant1	quant2
1	0.90703	.000003016	19.2468	5.09189

Example: A company is comparing methods for producing pipes and wants to choose the method with the least variability. It has taken a sample of the lengths of the pipes using both methods and found the following data and summaries. Find a 99% CI for the ratio of the variances. Values: $P(F_{11,14} > 4.508) = 0.005, P(F_{11,14} > 0.196) = 0.995, P(F_{14,11} > 5.103) = 0.005, P(F_{14,11} > 0.222) = 0.995$

The UNIVARIATE Procedure
Variable: Width

Method=A

Moments			
N	12	Sum Weights	12
Mean	4.06666667	Sum Observations	48.8
Std Deviation	0.88557463	Variance	0.78424242
Skewness	-0.0526034	Kurtosis	-1.6531257
Uncorrected SS	207.08	Corrected SS	8.62666667
Coeff Variation	21.7764253	Std Error Mean	0.25564338

Basic Statistical Measures			
Location		Variability	
Mean	4.066667	Std Deviation	0.88557
Median	4.050000	Variance	0.78424
Mode	3.100000	Range	2.50000
		Interquartile Range	1.70000

The UNIVARIATE Procedure
Variable: Width

Method=B

Moments			
N	15	Sum Weights	15
Mean	4.38	Sum Observations	65.7
Std Deviation	0.65159146	Variance	0.42457143
Skewness	0.15901561	Kurtosis	-1.1801064
Uncorrected SS	293.71	Corrected SS	5.944
Coeff Variation	14.8765173	Std Error Mean	0.16824019

Basic Statistical Measures			
Location		Variability	
Mean	4.380000	Std Deviation	0.65159
Median	4.400000	Variance	0.42457
Mode	3.500000	Range	2.00000
		Interquartile Range	1.20000

The hypothesis test for the ratio of the variances is summarized below:

Null	Alternative	Test Stat	RR
$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_0: \sigma_1^2/\sigma_2^2 \leq 1$	$H_A: \sigma_1^2 > \sigma_2^2$ $H_A: \sigma_1^2/\sigma_2^2 > 1$	S_1^2/S_2^2	$\{F_{obs} : F_{obs} \geq F_{\alpha,df1,df2}\}$
$H_0: \sigma_1^2 = \sigma_2^2$ $H_0: \sigma_1^2/\sigma_2^2 = 1$	$H_A: \sigma_1^2 \neq \sigma_2^2$ $H_A: \sigma_1^2/\sigma_2^2 \neq 1$	S_1^2/S_2^2	$\left\{F_{obs} : F_{obs} \geq F_{\alpha/2,df1,df2} \quad or \quad F_{obs} \leq F_{1-\alpha/2,df1,df2}\right\}$

Example: Recall heartrate example from chapter 6. Conduct an HT for equality of variance at the 0.05 level. $P(F_{14,10} > 3.798) = 0.025, P(F_{14,10} > 0.316)0.975$

The TTEST Procedure

Variable: rate

group	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	15	150.2	12.6500	3.2662	132.1	171.9
2	10	141.1	10.0004	3.1624	129.2	156.7
Diff (1-2)		9.1190	11.6849	4.7704		

group	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		150.2	143.2	157.2	12.6500	9.2614	19.9503
2		141.1	133.9	148.3	10.0004	6.8786	18.2568
Diff (1-2)	Pooled	9.1190	-0.7492	18.9872	11.6849	9.0817	16.3912
Diff (1-2)	Satterthwaite	9.1190	-0.3045	18.5425			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	23	1.91	0.0685
Satterthwaite	Unequal	22.202	2.01	0.0572

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	14	9	1.60	0.4833