Recap: When σ_1^2 and σ_2^2 are known

If the variances of both populations are **known**, then the hypotheses

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

may be tested using the statistic

Two-Sample Z-test

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{2.33}$$

If both populations are normal, or if the sample sizes are large enough so that the central limit theorem applies, the distribution of Z_0 is $N(0, 1)$ if the null hypothesis is true. Thus, the critical region would be found using the normal distribution rather than the t. Specifically, we would reject H_0 if $|Z_0| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution. This procedure is sometimes called the **two-sample *Z*-test**. A P-value approach can also be used with this test. The P-value would be found as $P = 2\,[1 - \Phi(|Z_0|)]$, where $\Phi(x)$ is the cumulative standard normal distribution evaluated at the point x.

The $100(1-\alpha)$ percent confidence interval on $\mu_1 - \mu_2$ where the variances are known is

$$\bar{y}_1 - \bar{y}_2 - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{y}_1 - \bar{y}_2 + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{2.34}$$

Summary: Tests IF Variance is Known

Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection	*P*-Value
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$\lvert Z_0 \rvert > Z_{\alpha/2}$	$P = 2[1 - \Phi(\lvert Z_0 \rvert)]$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$Z_0 = \dfrac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$	$Z_0 < -Z_\alpha$	$P = \Phi(Z_0)$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$Z_0 > Z_\alpha$	$P = 1 - \Phi(Z_0)$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$		$\lvert Z_0 \rvert > Z_{\alpha/2}$	$P = 2[1 - \Phi(\lvert Z_0 \rvert)]$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$Z_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$Z_0 < -Z_\alpha$	$P = \Phi(Z_0)$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$		$Z_0 > Z_\alpha$	$P = 1 - \Phi(Z_0)$

Summary: Tests IF Variance is Unknown

Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection	P-Value
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$\lvert t_0 \rvert > t_{\alpha/2,n-1}$	sum of the probability above t_0 and below $-t_0$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$t_0 = \dfrac{\bar{y} - \mu_0}{S/\sqrt{n}}$	$t_0 < -t_{\alpha,n-1}$	probability below t_0
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$t_0 > t_{\alpha,n-1}$	probability above t_0
	if $\sigma_1^2 = \sigma_2^2$		
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $v = n_1 + n_2 - 2$	$\lvert t_0 \rvert > t_{\alpha/2,v}$	sum of the probability above t_0 and below $-t_0$
	if $\sigma_1^2 \neq \sigma_2^2$		
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	$t_0 < -t_{\alpha,v}$	probability below t_0
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$v = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}$	$t_0 > t_{\alpha,v}$	probability above t_0

Example (DIY)

■ **TABLE 2.6**

Data for the Hardness Testing Experiment

Specimen	Tip 1	Tip 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5

Ref: Design and Analysis of Experiments, 8th Ed.

Consider a hardness testing machine that presses a rod with a pointed tip into a metal specimen with a known force. By measuring the depth of the depression caused by the tip, the hardness of the specimen is determined.

Two different tips are available for this machine, and although the precision (variability) of the measurements made by the two tips seems to be the same, it is suspected that one tip produces different mean hardness readings than the other. Is it so?

ME 794

Statistical Design of Experiments

Chapter 2.2

Classical Design of Experiments

Analysis of Variance (ANOVA)

Compare Variance with Fixed Value

Four observations on etch uniformity on silicon wafers are taken during a qualification experiment for a plasma etcher. The data are as follows:

6.00 7.25 5.25 6.50

CEP2022_Notebook (2.2)

Test the hypothesis that $\sigma^2 = 1.25$. Use $\alpha = 0.05$. Will you accept the hypotheses?

- What is the hypothesis test?
- Which statistic to use?
- Which test to use? (what to compare?)
- One-sided or Two-sided?

Remember Chi-Square (χ^2) Distribution?

If $z_1, z_2, z_3 \ldots, z_k$ are **normally and independently distributed** random variables with mean 0 and variance 1 [NID (0,1)]

And if we define, $x = z_1^2 + z_2^2 + \cdots + z_k^2$ Then 'x' follows the **chi-square distribution with k degrees of freedom**

The distribution is asymmetric or skewed: $\boldsymbol{\mu = k}$ $\boldsymbol{\sigma^2 = 2k}$

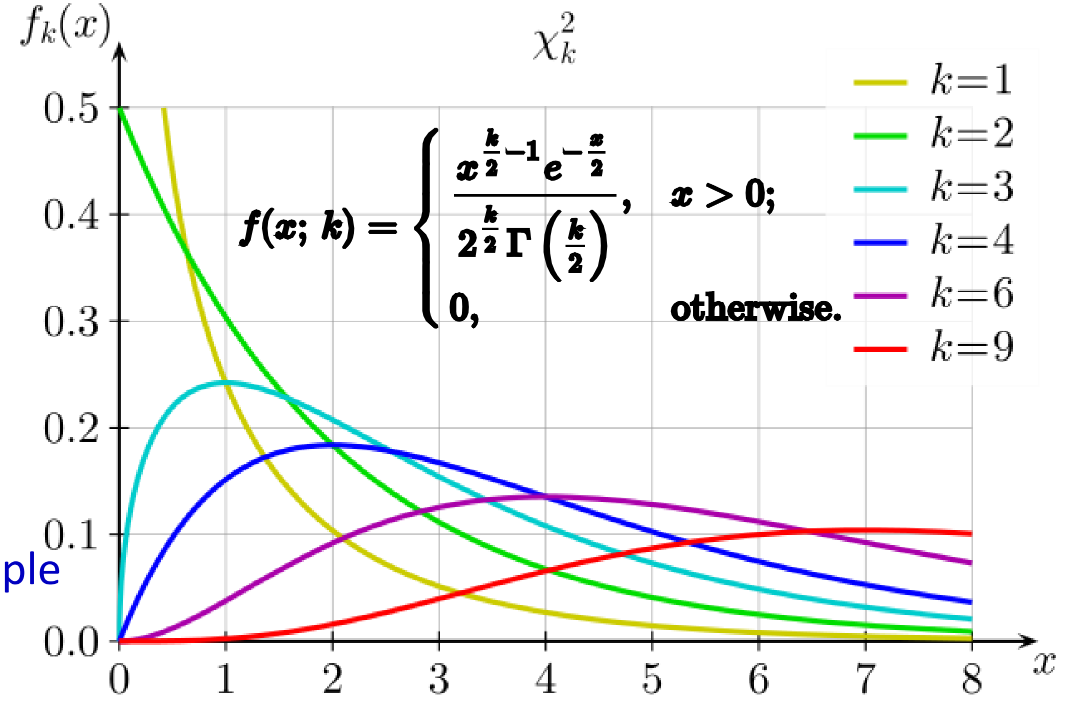
$SS = \sum_{i=1}^{n}(y_i - \bar{y})^2$ is the **corrected sum of squares** of the observations y_i.

$$E(S^2) = \frac{1}{n-1}E(SS) = \sigma^2$$

and we see that S^2 is an unbiased estimator of σ^2.

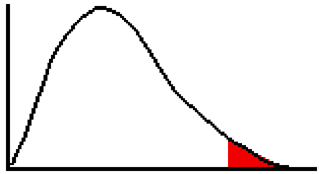
$$\frac{SS}{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Sample variance, $S^2 = \frac{SS}{n-1}$ Therefore, if the observations in the sample are NID (μ, σ^2), then the distribution of S^2 *is* $\left(\frac{\sigma^2}{n-1}\right)\chi^2_{n-1}$

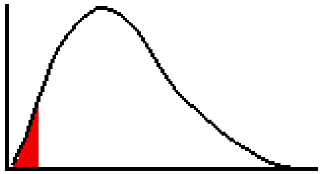


Chi-Squared Table

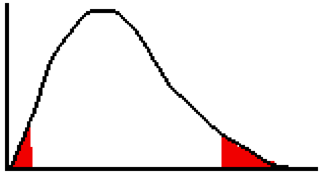
HOW TO USE THIS GRAPH:



To find this region use the value equivalent to α at the top of the table.



To find this region use the value equivalent to $1-\alpha$ at the top of the table.



To find the region to the left, use $1-\alpha/2$.
To find the region to the right, use $\alpha/2$.

degrees of freedom	Area to the right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	-----	-----	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.383
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.544	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997

What if we want to compare > 2 samples?

- In the Portland cement experiment, two different mortar formulations were tested. This experiment can be considered as an experiment with one-factor at two levels.
- We used the **two-sample t-test** along with the confidence interval approach for comparing two techniques/methods/products.
- Several experiments would involve more than two levels of the factors.
- **We now want to extend the analysis to consider more than two technique comparisons -> using Analysis Of Variance (ANOVA)**

■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Example

Experiments were carried out to determine the corrosion rates of four different metals. Specimens of each of FOUR different metals were immersed in a highly corrosive solution, and these corrosion rates in percentages were recorded. A simple question to be asked is: **"Are the corrosion rates different for different metals?"** OR **'is there evidence to indicate any real differences among the mean values associated with different metal corrosion rates?'**

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59
	μ_1	μ_2	μ_3	μ_4

Hypothesis Testing

Null Hypothesis: H_0

$$\mu_1 = \mu_2 = \mu_3 = \mu_4$$

Alternate Hypothesis: H_1

$$\mu_1 \neq \mu_2 \neq \cancel{\mu_3} \neq \mu_4$$

At least one μ_i is diff from rest

ANOVA: Mathematical Model

- **(Remember)** What was our model for data with a single variable?

"Means Model"

$$y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \ldots, n_i \end{cases}$$

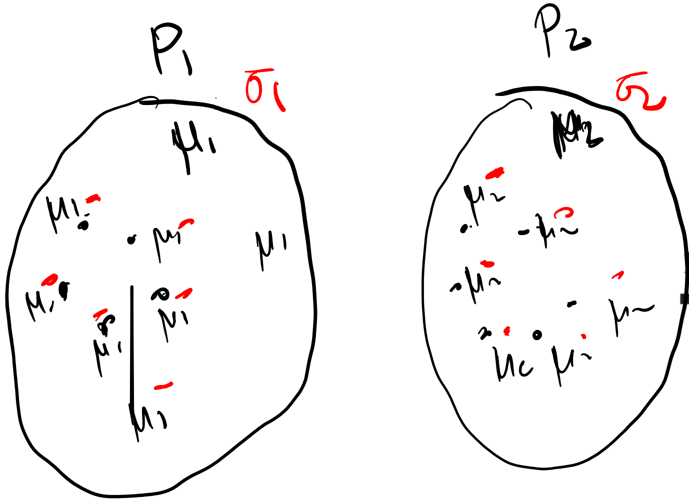
where y_{ij} is the jth observation from factor level i, μ_i is the mean of the response at the ith factor level, and ϵ_{ij} is a normal random variable associated with the ijth observation.

OR

$$y_{ij} = \bar{y}_i + (y_{ij} - \bar{y}_i)$$

Observed Value = Mean Value + Error

$$\underbrace{y_{ij}}_{\text{obs v}} = \bar{y}_i + \underbrace{(y_{ij} - \bar{y}_i)}_{\epsilon_{ij}}$$



■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

$\mu_1 = \mu_2$

$\mu_1 \neq \mu_2$

ANOVA: Mathematical Model

all data mean $\bar{\bar{y}}$

What is our mathematical model when we have multiple sets of data from multiple tests? i.e., multiple levels (>2) of a variable

"Effects Model"

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i)$$

Obs. value | Grand mean | Deviation of treatment mean from Grand Mean | Deviation of Obs. value from treatment mean [Intrinsic Variation]

j	Stainless Steel (y_{1j})	Carbon Steel (y_{2j})	Nickel Alloy (y_{3j})	High Speed Steel (y_{4j})
1	62	63	68	56 y_{41}
2	60	67	66	62
3	63 y_{13}	71	71 y_{33}	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66 y_{26}	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59
	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$

y_{ij}

jth observation for ith metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

"effects model"

⇒ if H_1 : Effect is true ⇒ obs value = Grand mean + Dev. of treatment from Grand mean + Residual

if H_0 : Effect absent ⇒ obs value = Grand mean + Residual* → "Means model"

ANOVA: Mathematical Model

$k=$	1 $\longrightarrow$	2	3	4
j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59
	$n_1=4$	$n_2=6$	$n_3=6$	$n_4=8$

Effects Model

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i) \quad \text{—— (1)}$$

take sq. of both sides and sum over all i & j

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\Big[\underbrace{\bar{\bar{y}}}_{a} + \underbrace{(\bar{y}_i - \bar{\bar{y}})}_{b} + \underbrace{(y_{ij} - \bar{y}_i)}_{c}\Big]^2$$

$$= \sum\sum\Big[\bar{\bar{y}}^2 + (\bar{y}_i - \bar{\bar{y}})^2 + (y_{ij} - \bar{y}_i)^2 + 2\bar{\bar{y}}\underbrace{(\bar{y}_i - \bar{\bar{y}})}_{0} + 2\bar{\bar{y}}\underbrace{(y_{ij} - \bar{y}_i)}_{0} + 2\underbrace{(\bar{y}_i - \bar{\bar{y}})(y_{ij} - \bar{y}_i)}_{0}\Big]$$

$$\sum\sum y_{ij}^2 = \sum\sum\Big[\bar{\bar{y}}^2 + (\bar{y}_i - \bar{\bar{y}})^2 + (y_{ij} - \bar{y}_i)^2\Big] \quad \text{—— (2)}$$

y_{ij}

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

$\rightarrow n_i$ = sample size for i^{th} metal

ANOVA: Mathematical Model

$\sum n_i(\bar{y}_i - \bar{\bar{y}}) = 0$

$$\sum_{i}^{k}\sum_{j}^{n_i} y_{ij}^2 = \sum_{i}^{k}\sum_{j}^{n_i}\left[\bar{\bar{y}}^2 + (\bar{y}_i - \bar{\bar{y}})^2 + (y_{ij} - \bar{y})^2\right]$$

$$\underbrace{\sum\sum y_{ij}^2}_{SS_{total}} = \underbrace{N\bar{\bar{y}}^2}_{SS_{mean}} + \underbrace{\sum_{i=1}^{k} n_i(\bar{y}_i - \bar{\bar{y}})^2}_{SS_{treatment}} + \underbrace{\sum_{i}^{k}\sum_{j}^{n_i}(y_{ij} - \bar{y}_i)^2}_{SS_{intrinsic}}$$

"Grand Mean"

"Between treatments"

or SS_{error}

or "Intrinsic Variation"

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

y_{ij} $\quad N = 24$

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

$N = \sum_{i}^{k} n_i$ = total # of observations

$$\boxed{SS_{total} = SS_{mean} + SS_{treatment} + SS_{Error}}$$

"Effects Model" (H_1)

ANOVA: Mathematical Model

$\sum\sum y_{ij} - \bar{\bar{y}} = 0$

64

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

y_{ij}

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

In some textbooks, they use $SS_{c.total}$

$=$ 'Corrected' total SS $= \sum_{i}^{k}\sum_{j}^{n_i}(y_{ij} - \bar{\bar{y}})^2$

In that case,

$SS_{c.total} = SS_{treatment} + SS_{error}$

Note that, $SS_{c.total} = SS_{total} - SS_{mean}$

i.e., $\sum_{i}^{k}\sum_{j}^{n_i}(y_{ij} - \bar{\bar{y}})^2 = \sum_{i}^{k}\sum_{j}^{n_i} y_{ij}^2 - N\bar{\bar{y}}^2$ [DIY]

$SS_{total} = SS_{mean} + SS_{treatment} + SS_{error}$

DOF $\Rightarrow$ N, 1, k − 1, N − k

$SS_{c.total}$ has N−1 dof

ANOVA: Mathematical Model

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i)$$

$$y_{ij} = \underbrace{\mu}_{\text{pop mean}} + \underbrace{\tau_i}_{\text{effect of treatment}} + \underbrace{\varepsilon_{ij}}_{\text{residual error}}$$

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

y_{ij}

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

$$E(y_{ij}) = E(\mu + \tau_i + \varepsilon_{ij}) = E(\mu) + E(\tau_i) + \underbrace{E(\varepsilon_{ij})}_{0}$$

$$= \mu + \tau_i + 0$$

Because errors are NID $(0, \sigma^2)$

$$Var(y_{ij}) = Var(\mu + \tau_i + \varepsilon_{ij}) = Var(\mu) + Var(\tau_i) + Var(\varepsilon_{ij})$$

$$= 0 + 0 + \sigma^2$$

Thus, y_{ij} follows a normal distribution with mean $= \mu + \tau_i$ and variance $= \sigma^2$

ANOVA: Mathematical Model

Therefore, $\frac{SS}{\sigma^2}$ of y_{ij} follows χ^2_N distribution

$$\frac{SS_{treatment}}{\sigma^2} = \frac{\sum^{k} n_i(\bar{y}_i - \bar{\bar{y}})^2}{\sigma^2} \rightarrow \chi^2_{k-1}$$

$$\frac{SS_{error}}{\sigma^2} = \frac{\sum^{k}\sum^{n_i}(y_{ij} - \bar{y}_i)^2}{\sigma^2} \rightarrow \chi^2_{N-k}$$

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

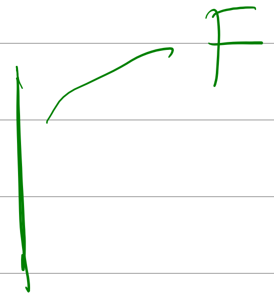
y_{ij}

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

$$\frac{\dfrac{SS_{treatmen}}{\cancel{\sigma^2}(k-1)}}{\dfrac{SS_{error}}{\cancel{\sigma^2}(N-k)}} = \frac{\chi^2_{k-1}/k-1}{\chi^2_{N-k}/N-k}$$



ANOVA

Justification for using F-distribution for hypothesis testing

- Cochran proved a theorem that these chi-square distributions will be independent if the total DOF is equal to the sum of the other sum-of-squares DOF.
- In our case, the DOF for $SS_{treatment}$, SS_{error} and SS_{mean} add to the total DOF, then these sum of squares can be considered independent.

$MS = \frac{SS}{\nu}$

Therefore, under the Null Hypothesis, the statistic,

$$F_0 = \frac{\left(\frac{SS_{treatment}}{k-1}\right)}{\left(\frac{SS_{error}}{N-k}\right)} = \frac{MS_{treatment}}{MS_{error}}$$

needs to be large?

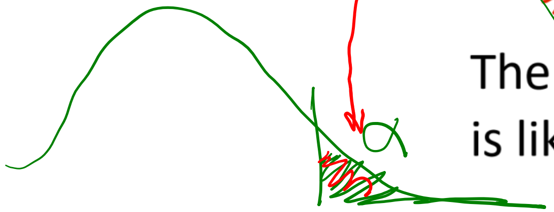
$> F_{1-\alpha, \nu_1, \nu_2}$

follows F-distribution with (k-1, N-k) DOF

One sided?

$SS_{treatment} >> SS_{error}$

for H_1 to be true



If $F_{cal} = \frac{S_1^2}{S_2^2} >$ **Tabulated value of** $F(\nu_1, \nu_2)$

$k-1$

$N-k$

Then there is significant difference in mean squares is likely.