Quantitative Methods

Lecture-11

BITS Pilani
Pilani Campus

innovate achieve lead

BITS Pilani
Pilani Campus

Two Sample Tests and ANOVA

(Ch 10 Business Statistics, Levine et al.)

So far and the next

- ✓ Previous Sessions
- √ Probability Distributions
- √ Confidence Interval Estimation
- √ Hypothesis testing
- ➢ Today
  - ➢ Two Sample Tests and ANOVA

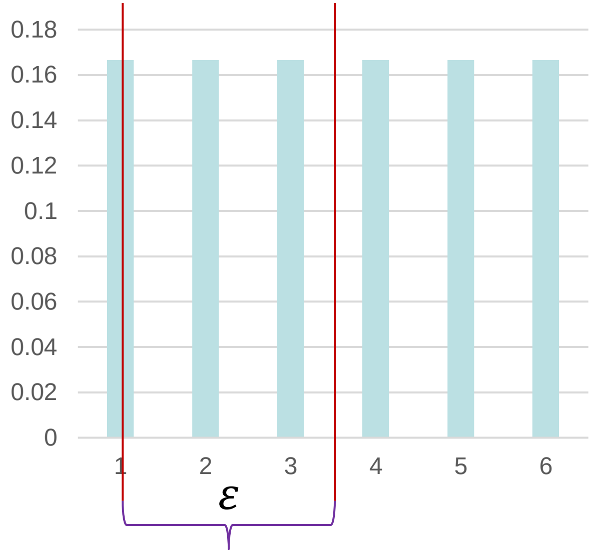
Concept Alert! Inherent randomness

<u>Population</u>

- All outcomes of a random variable (X) under consideration.
- Population distributions can be discrete (binomial etc.) or continuous (normal etc.).

<u>A fair dice population</u>

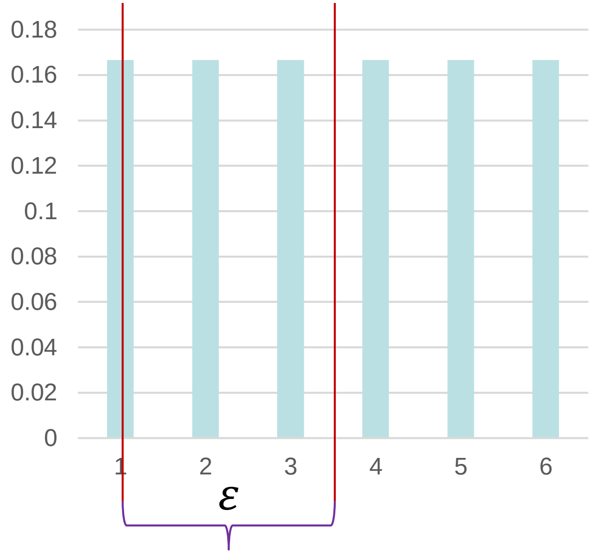
- Random Var. X: The number that shows up on the top.
- The population: The distribution of for example, a billion outcomes of X.
- μ ~3.5, σ ~ 1.71.



$$X_i = \mu + \varepsilon_i$$

Concept Alert! Inherent randomness

X: μ ~3.5, σ ~ 1.71.



$$X_i = \mu + \varepsilon_i$$

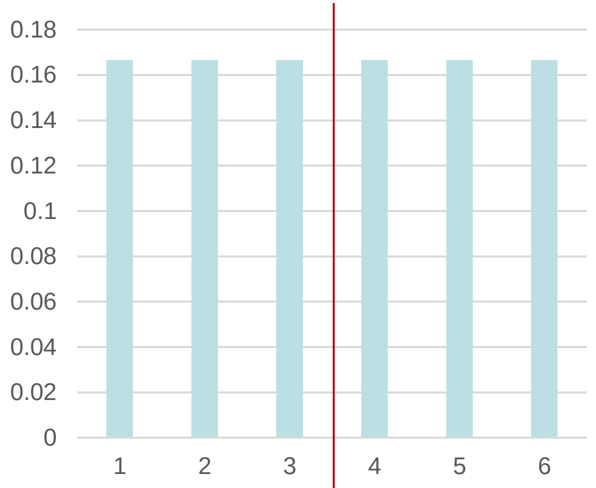
<u>Sampling error</u>

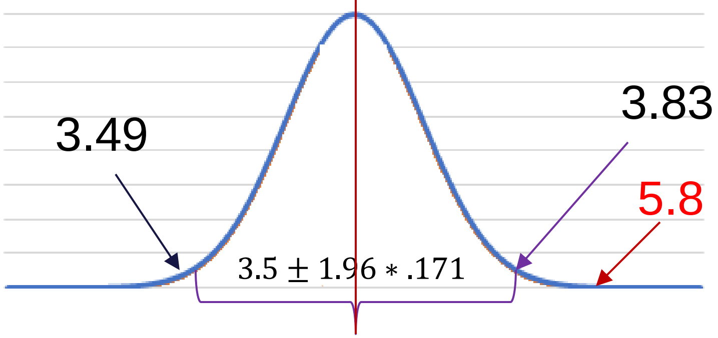
- Say you draw one random sample of one of the outcomes, X_i.
- Will the value of X_i be 3.5?
- Every time you throw this dice, you will get: $X_i = \mu + \varepsilon_i$.
- Mean + some random error (ε_i).
- This random error (ε_i) is because of **inherent randomness** / **sampling error / random error.**

Concept Alert! Same or different populations?

<u>Sampling error?</u>

- From the population X, you draw a random sample of 100 outcomes.
- Every time you draw a sample and calculate the mean.
- You would get $\overline{Xi} = \mu + \varepsilon_{\text{i}}$ – Mean + some random error
- You want to find out the possible reason of the difference between μ & $\overline{Xi}$
- Is it because of inherent randomness or sampling error?
- Or, does the sample belong to a different population?
  - The difference cannot be explained by inherent randomness.



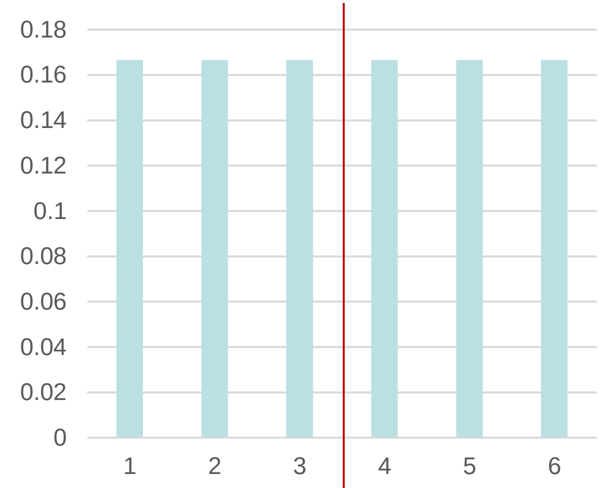


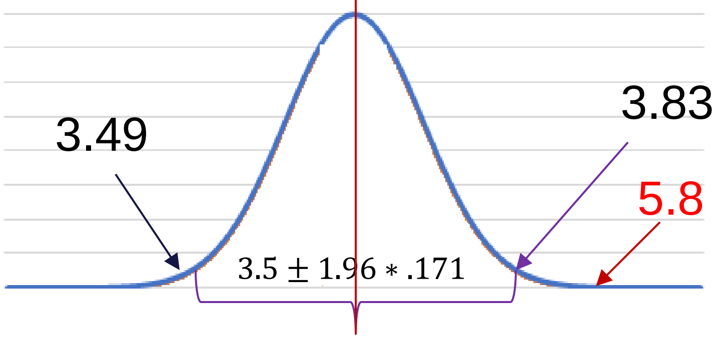
n=100, Sampling distribution of $\overline{X}$. $\sigma_{\bar{X}} = 0.171$

Concept Alert! Same or different populations?

<u>Questions we generally would be answering</u>

- We would have some proven sampling distribution for our *sample statistic*
- We would ask: Is the sample statistic within the region of "non-rejection".
- We assume that our sample statistic can be anywhere in the region of non-rejection because of the inherent randomness.
- That is the core of all hypothesis testing.
- For example, if you found the mean of 100 throws of dice to be 5.8. Does it belong to fair dice population?
- You can say with 95% confidence that the mean is too far to belong to the same population





n=100, Sampling distribution of $\overline{X}$. $\sigma_{\bar{X}} = 0.171$

Practical Applications

- Is there really a difference beyond just plain randomness?
- Are results from two marketing campaigns same?
- Are prices from two stores same?
- Are the conversion rates the same, for the "buy-now button" of 5 different colors on your website?
- Is stock portfolio-A riskier (more variance) than the portfolio-B?
- **Whether there is a statistical difference between the means, proportions, standard deviations etc.?**

Comparing means of two populations

Comparing the means of two independent populations

Comparing the means of two populations

- Is there any difference between two population means?(H_0?)
- H_0: μ1 = μ2.
- Can also be written as: H_0: μ1 - μ2 = 0
- H_1: μ1 ≠ μ2

Independent populations?

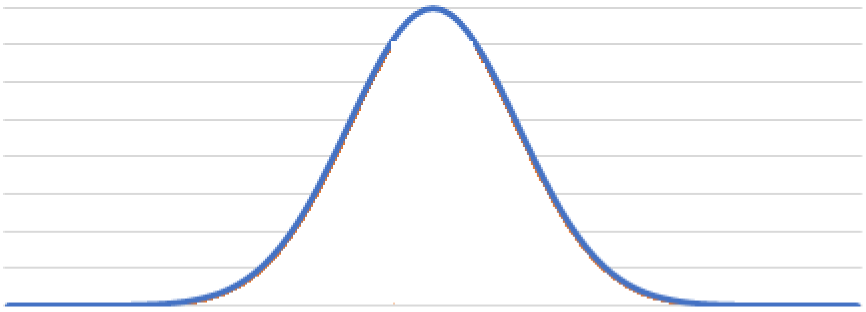
- Units in one are not related to units in another (Members of a family).

Assumptions

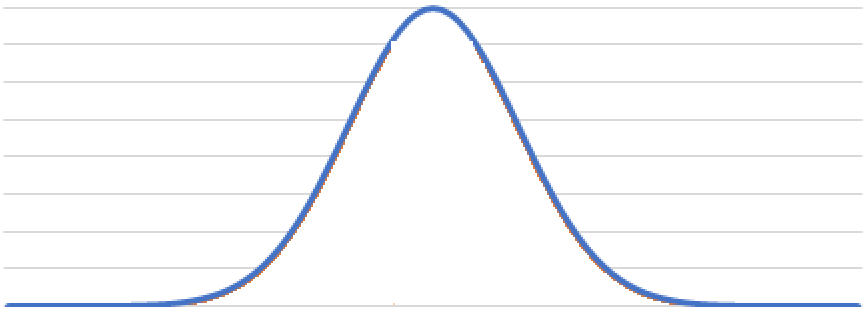
- t-test: Normality of the populations (when σ is not known).
- Equal variance.

Difference in Means: Sampling Distribution

$X1 \sim N(\mu1, \sigma^2)$

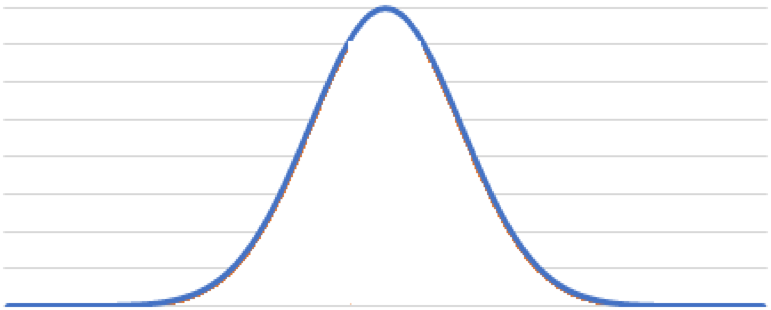


$X2 \sim N(\mu2, \sigma^2)$



Populations

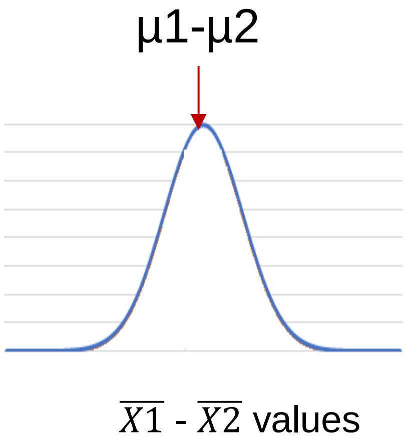
$\overline{X1} - \overline{X2} \sim N(\mu1-\mu2, \sigma_{\overline{X1} - \overline{X2}}^2)$



$\overline{X1} - \overline{X2}$ values

Sampling Distribution of Differences in Means

Difference in Means: Sampling Distribution



Standard error:

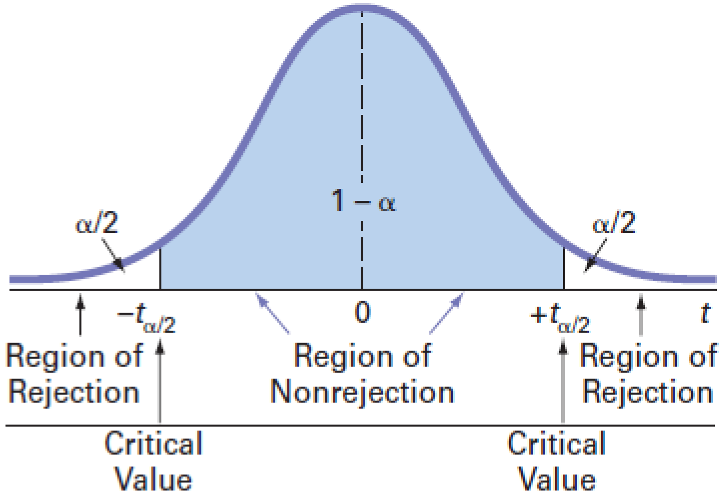
$$Standard\ error:\ \sigma_{\overline{X1}\,-\,\overline{X2}}{}^{2} = \sqrt{Sp^2\left(\frac{1}{n1}+\frac{1}{n2}\right)}$$

- Sp^2: Pooled common variance estimate of the populations
- **Why pooled?**
- We have two estimates, pooling them gives an estimate from a larger sample
- $S_p{}^2 = \frac{(n_1-1)*s_1{}^2 + (n_2-1)*s_2{}^2}{n_1+n_2-2}$
- **Confidence interval**
- $(\overline{X1} - \overline{X2}) \pm t_{\frac{\alpha}{2}}\sqrt{Sp^2\left(\frac{1}{n1}+\frac{1}{n2}\right)}$
- $t_{\frac{\alpha}{2}}$: Upper tail critical value, with α significance, of t distribution with n_1+n_2 -2 degrees of freedom (critical value for upper tail area of $\frac{\alpha}{2}$)

Hypothesis testing:

<u>Independent populations: Comparison of means</u>

- Are population means same, one greater or lower?
- Two tailed: H_0: μ1 = μ2 => μ1 - μ2 = 0
- One tail: H_0: μ1 ≥ μ2 => μ1 - μ2 ≥ 0



<u>Pooled Variance t-test</u>

- Find critical t-values that separate region of non-rejection and region of rejections
- $t_{\frac{\alpha}{2}}$: Upper tail critical value, with α significance, with n_1+n_2 -2 degrees of freedom
- Decide whether the sample t_{STAT} falls in the region of rejection
- Alternatively, find p-value for the t_{STAT} and compare it with the level of significance
- $t_{STAT} = \frac{(\overline{X1} - \overline{X2}) - (\mu 1 - \mu 2))}{\sqrt{Sp^2(\frac{1}{n1}+\frac{1}{n2})}}$, where $S_p{}^2 = \frac{(n_1-1) * S_1{}^2 + (n_2-1) * S_2{}^2}{n_1+n_2-2}$

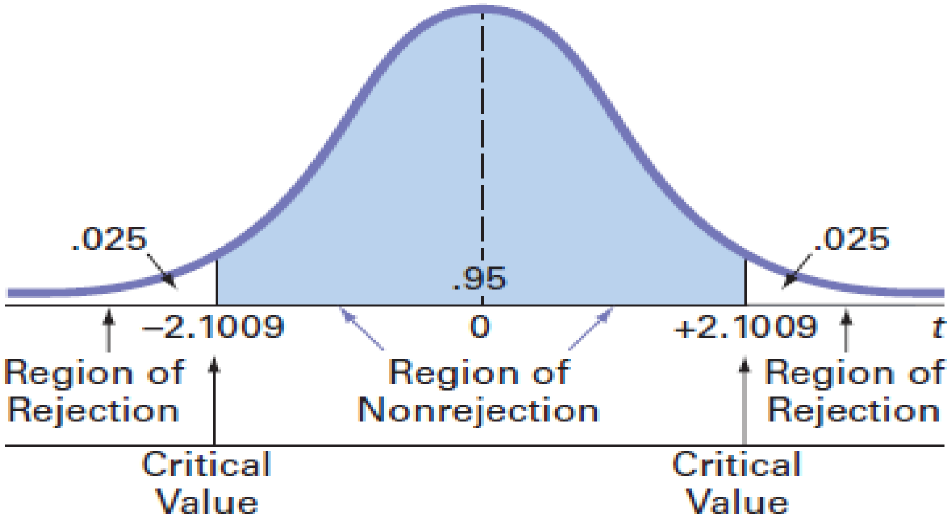
Is there a difference in store sales?

You have two stores that sell your coconut water.
You want to test if the sale at these two locations is the same or different at 0.05 level of significance.
You draw 10 samples from each store sales data.
You find that the samples means are 50.3 and 72 bottles resp, and
The Sample standard deviations were found to be 18.73 and 12.54 resp.

- H_0 and H_1
- $\mu1 = \mu2$ => $\mu1 - \mu2 = 0$ and H_1: $\mu1 \neq \mu2$ => $\mu1 - \mu2 \neq 0$
- Test statistic: $\overline{X1} - \overline{X2}$?
- -22
- Distribution of the test statistic (if null was true)?
- t-distribution with difference of mean to be zero.
- Degrees of freedom for test statistic?
- n1 + n2 -2 = 18

Example

- Test statistic: $\overline{X1} - \overline{X2}$: -22
- Standard error?
- Pooled variance: $S_p{}^2 = \frac{(n_1-1)*S_1{}^2 + (n_2-1)*S_2{}^2}{n_1+n_2-2}$ = 254
- Standard error: $\sqrt{Sp^2(\frac{1}{n1} + \frac{1}{n2})}$ = 7.13
- Degrees of freedom: 18
- $t_{STAT} = \frac{(\overline{X1} - \overline{X2}) - (\mu 1 - \mu 2))}{\sqrt{Sp^2(\frac{1}{n1} + \frac{1}{n2})}}$ = -3.044
- **Critical values: -2.1 and +2.1 , p-value: 0.007**
- p-value < α and t_{STAT} is in the region of rejection.



Comparing the means of two related populations

<u>When two populations (for sample 1 and sample 2) are not independent</u>

- Outcomes of the first population are not independent of the outcomes of the second population

- Repeat measurements on same units:

Before	**After**	**Difference**
X1b	X1a	D1 = X1a - X1b
X2b	X2a	D2 = X2a – X2b

- You treat differences (D1, D2, …, Dn) as population
- Draw a random sample from that.
- Null hypothesis?
- H_o: $\mu_D = 0$ and H_1: $\mu_D \neq 0$.
- t_{STAT} for (n-1 degrees of freedom):

$(\overline{D} - \mu D))/S_D/\sqrt{n}$.

Comparing the means of two related populations

<u>When two populations (for sample 1 and sample 2) are not independent</u>

- Are store-A book prices more than store-B prices?
- If the sample “A” predominantly consists of very expensive titles and B contained predominantly inexpensive paperbacks.
- Would the results be valid?
- Better way may be to take the *prices for similar books for comparison.*
- Another way is to pair the measurement units on some characteristic (paperbacks and hardbound etc)

Store-A	**Store-B**	**Difference**
Title1	Title1	D1 = T1a - T1b
Title2	Title2	D2 = T2a – T2b

Why use related populations?

- Samples may not be comparable
- Using unrelated items may make variance very high.
- By using related populations, we reduce overall variance by reducing the “*unit level difference*” between the two samples.
- Lower variance reduces the β-error (probability of not rejecting a false null hypothesis).

innovate achieve lead

BITS Pilani
Pilani Campus

Comparing proportions of two populations

Comparing the proportions of two independent populations

- Are population proportions same?

Two tailed:

H_0: π1=π2

$\Rightarrow$ π1 – π2 = 0

H_1: π1≠π2 => π1 – π2 ≠ 0

One tail:

H_0: π1 $\geq$ π2

$\Rightarrow$ π1 – π2 $\geq$ 0

H_1: π1 $< or >$ π2

=> π1 – π2 < or > 0

- Z statistic approximately follows Z-Distribution

Comparing the proportions of two independent populations

Standard error:

$$Standard\ error:\ \sigma_{p1-\ p2} = \sqrt{\bar{p}(1-\bar{p})\ (\frac{1}{n1}+\frac{1}{n2})}$$

- **Pooled proportion estimate:**
- $\bar{p} = \frac{X1+X2}{n1+n2},\ \text{p1} = \frac{X1}{n1},\ \text{p2} = \frac{X2}{n2}$
- Confidence interval
- $(p1-\ p2) \pm Z_{\alpha/2}\sqrt{\bar{p}(1-\bar{p})\ (\frac{1}{n1}+\frac{1}{n2})}$

$$Z_{STAT} = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

p_1 = proportion of items of interest in sample 1
X_1 = number of items of interest in sample 1
n_1 = sample size of sample 1
π_1 = proportion of items of interest in population 1
p_2 = proportion of items of interest in sample 2
X_2 = number of items of interest in sample 2
n_2 = sample size of sample 2
π_2 = proportion of items of interest in population 2
$\bar{p}$ = pooled estimate of the population proportion of items of interest

All the steps are similar to the previous example. The only difference is that we now use a Z-test instead of a t-test.

innovate achieve lead

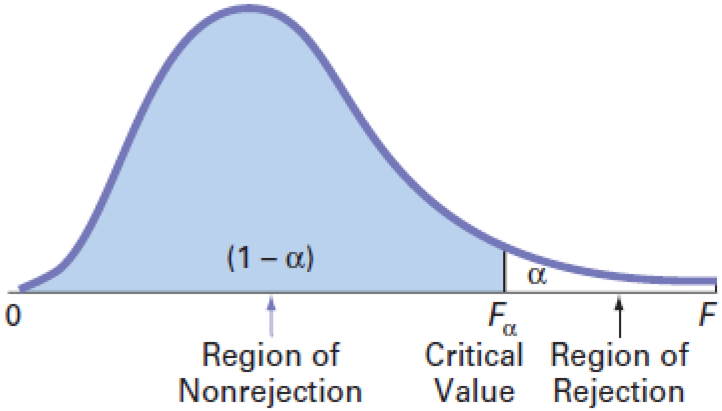
BITS Pilani
Pilani Campus

Comparing variances of two populations

Comparing variances of two independent populations

Testing for equality of variances

- H_o: $\sigma1^2 = \sigma2^2$ or $\frac{\sigma1^2}{\sigma2^2} = 1$ and
- H_1: $\sigma1^2 \neq \sigma2^2$



- F-Test
- If two independent populations are normally distributed.
- Ratio of sample variances follow F-Distribution.
- $F_{STAT} = S_1^2 / S_2^2$
- Follows F-Distribution with
- *n1-1* numerator degrees of freedom and *n2-1* denominator degrees of freedom
- Larger variance is taken as numerator.
- Critical F-Value is calculated from the F-Table.
- Reject null if $F_{STAT} > F_{\alpha/2}$ or if p-value < α.

Note:
Pic depicts one tail F-test
Though the left side discussion is about one tail test

Equality of variances, two independent populations

- You have two stores that sell your coconut water.
- You want to test if the sale at these two locations is same or different at 0.05 level of significance.
- You draw 10 samples (10 days) from each store
- You find that the mean sale volume to be 50.3 and 72 bottles resp.
- Sample standard deviations were found to be 18.73 and 12.54 resp.
- **One of the important assumption of pooled t-test is that population variances are same**
- **We will now test that assumption!**
- H_o: $\sigma 1^2 = \sigma 2^2$ or $\frac{\sigma 1^2}{\sigma 2^2} = 1$
- H_1: $\sigma 1^2 \neq \sigma 2^2$

Equality of variances, two independent populations

F-Test

$F_{STAT} = S_1^2 / S_2^2$

350.68/157.33 = 2.23

Numerator & Denominator degrees of freedom:

9

$F_{critical}$? (From F-Table for .05 significance)

4.03

Should we reject the null?

$F_{STAT} < F_{critical}$

p-value = 0.25 ($p>\alpha$)

We do not have sufficient evidence to reject the null hypothesis.

We can assume that the variances are equal.

Cumulative Probabilities = 0.975
Upper-Tail Area = 0.025
Numerator df_1

Denominator df_2	1	2	3	...	7	8	$\boxed{9}$
1	647.80	799.50	864.20	...	948.20	956.70	963.30
2	38.51	39.00	39.17	...	39.36	39.37	39.39
3	17.44	16.04	15.44	...	14.62	14.54	14.47
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
7	8.07	6.54	5.89	...	4.99	4.90	4.82
8	7.57	6.06	5.42	...	4.53	4.43	4.36
$\boxed{9}$	7.21	5.71	5.08	...	4.20	4.10	$\boxed{4.03}$

Source: Extracted from Table E.5.

ANOVA: Analysis of Variance

ANOVA: Analysis of variance

- So far, we learned about comparing two populations (means and variances).
- ANOVA helps compare means of more than two populations or groups.
- The criteria that separate the groups are called Factors.
- Each factor can have multiple levels (or categories).

<u>One-way ANOVA:</u>

- One factor
- Levels separate the groups
- Also known as completely randomized design

ANOVA: Analysis of variance

It’s a two Step Process

- Find if there is difference population means (Ho: μ1 = μ2 = μ3 = μ4), H_1?
- H1: Not all means are same or at least one mean is different from others
- If null is rejected, find out which means are different from others
- Second step is not in the syllabus, but we may review it if time permits.

ANOVA: Analysis of variance

- Assuming you randomly select samples from “c” populations, which are normally distributed with equal variances
- Null hypothesis: Ho: $\mu_1 = \mu_2 = \dots = \mu_c$ (Means of all populations are equal)
- Alternate hypothesis: Not all μ_js are equal (Where, j=1,2,…,c).
  - At least one population mean is different from the other population means
- We are going to use the division of variations for testing this hypothesis.

	Loc 1	Loc 2	Loc3	Loc4
	30.06	32.22	30.78	30.33
	29.96	31.47	30.91	30.29
	30.19	32.13	30.79	30.25
	29.96	31.86	30.95	30.25
	29.74	32.29	31.13	30.55
Mean	**29.982**	**31.994**	**30.912**	**30.334**
SD	**0.164985**	**0.335306**	**0.142548**	**0.12522**

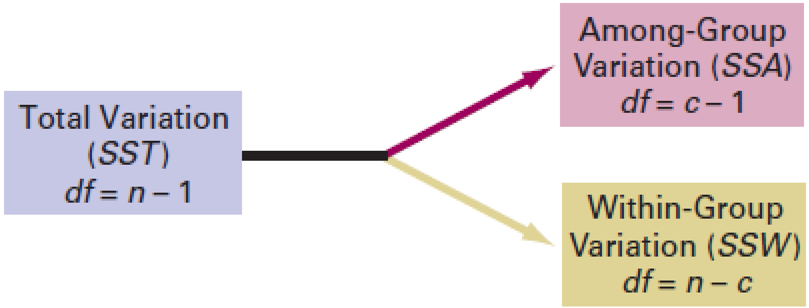
ANOVA: Analysis of variance

We partition the total variation into two parts.

- Within group variation: Due to random variation or the sampling error (SSW).
  - Differences are due to the sampling error or are random ($X_i = \mu + \varepsilon_i$).
- Among groups variation: Due to variation between the groups (SSA)

- Total Variation of the joint sample (SST) = SSW + SSA

	Loc 1	Loc 2	Loc3	Loc4
	30.06	32.22	30.78	30.33
	29.96	31.47	30.91	30.29
	30.19	32.13	30.79	30.25
	29.96	31.86	30.95	30.25
	29.74	32.29	31.13	30.55
Mean	**29.982**	**31.994**	**30.912**	**30.334**
SD	**0.164985**	**0.335306**	**0.142548**	**0.12522**



ANOVA: Total Sum of Squares

Calculating the total sum of squares (Total variation): SST

- Under null hypothesis, all the group means are same, so they all can just be considered to be coming from one population
- We calculate grand mean $\bar{\bar{X}}$ and then calculate sum of square deviation from all the values
- SST = $\sum_{j=1}^{c}\sum_{i=1}^{n_i}(X_{ij}-\bar{\bar{X}})^2$, ($X_{ij}$: ith value from jth group)
- Grand mean = $\bar{\bar{X}} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n_i}(X_{ij})}{n}$
- n = Total sample size (sum of all group sample sizes)
- Associated degrees of freedom with the SST?
- n-1
- Total Sum of Square Variance for ANOVA
  - Total Mean Squares: MST = SST / (n-1)

	Loc 1	Loc 2	Loc3	Loc4
	30.06	32.22	30.78	30.33
	29.96	31.47	30.91	30.29
	30.19	32.13	30.79	30.25
	29.96	31.86	30.95	30.25
	29.74	32.29	31.13	30.55
Mean	**29.982**	**31.994**	**30.912**	**30.334**
SD	**0.164985**	**0.335306**	**0.142548**	**0.12522**

Among Group Variations

<u>Calculating the among group sum of squares : SSA</u>

- Sum of squares of deviation of each mean from the grand mean
- Weighted by the group sample size (each group sample size may be different)
- SSA = $\sum_{j=1}^{c} n_j(\bar{X}_j - \bar{\bar{X}})^2$
- Associated degrees of freedom with the SSA : c-1
- Among group Variance for ANOVA
  - Mean Squares Total: MSA = SSA / (c-1)

	Loc 1	Loc 2	Loc3	Loc4
	30.06	32.22	30.78	30.33
	29.96	31.47	30.91	30.29
	30.19	32.13	30.79	30.25
	29.96	31.86	30.95	30.25
	29.74	32.29	31.13	30.55
Mean	**29.982**	**31.994**	**30.912**	**30.334**
SD	**0.164985**	**0.335306**	**0.142548**	**0.12522**

Within Group Variations

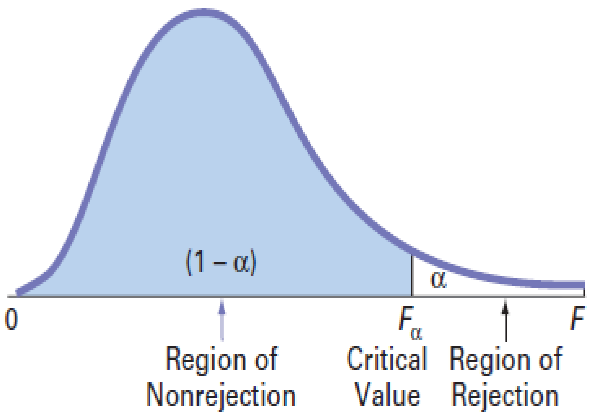
Calculating within groups sum of squares : SSW

- Sum of squares of individual value from their group means
- SSW = $\sum_{j=1}^{c} \sum_{i=1}^{n_i} (X_{ij} - \bar{X}_j)^2$
- Associated degrees of freedom with the SSW: n-c
- (Each group contributes to n-1 degrees of freedom)
- Within group Variance for ANOVA
  - Mean Squares Total: MSA = SSA / (n-c)

	Loc 1	Loc 2	Loc3	Loc4
	30.06	32.22	30.78	30.33
	29.96	31.47	30.91	30.29
	30.19	32.13	30.79	30.25
	29.96	31.86	30.95	30.25
	29.74	32.29	31.13	30.55
Mean	**29.982**	**31.994**	**30.912**	**30.334**
SD	**0.164985**	**0.335306**	**0.142548**	**0.12522**

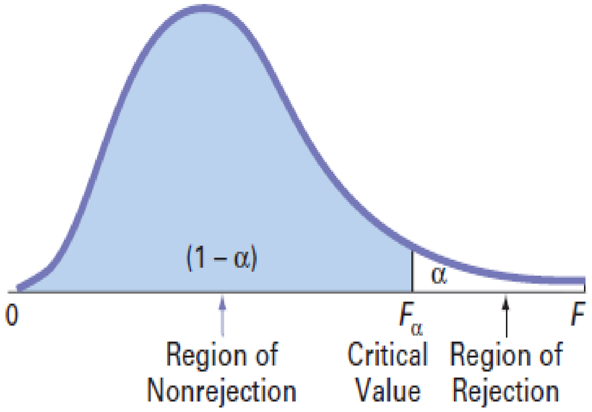
Using F-Test for One Way ANOVA s

- We used F-Test to test the equality of two variances
- Under the null hypothesis all groups belong to same population.
- Given this, grouped variance estimate (MSW) and within variance estimate (MSW) should be equal.
- One-Way ANOVA test statistic: $F_{STAT} = \frac{MSA}{MSW}$
- Follows F-Distribution with, *c-1* numerator degrees of freedom, and *n-c* denominator degrees of freedom



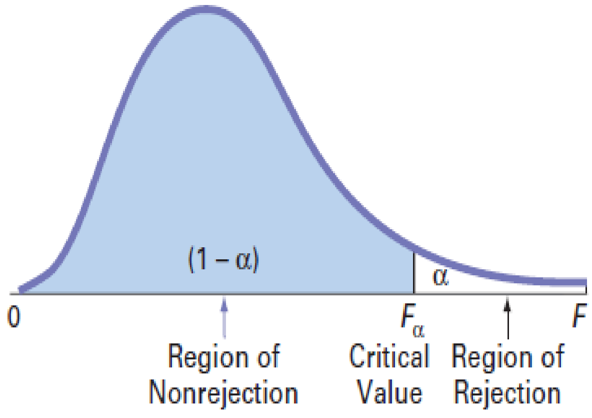
Using F-Test for One Way ANOVA s

- Reject null if $F_{STAT} > F_{\alpha}$
- Why not $F_{\alpha/2}$?
- MSW is only capturing the sampling error: SSW = $\sum_{j=1}^{c} \sum_{i=1}^{n_i} (X_{ij} - \bar{X}_j)^2$
- MSA additionally captures potential group differences if exist.
  - SSA = $\sum_{j=1}^{c} n_j (\bar{X}_j - \bar{\bar{X}})^2$
- If all means are not equal. MSA> MSW
- <u>This concludes the first step.</u>
- Rejection of null hypothesis only means that not all means are equal.
- If null was rejected, we still do not know - which one?



ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Square (Variance)	F
Among groups	$c - 1$	SSA	$MSA = \dfrac{SSA}{c-1}$	$F_{STAT} = \dfrac{MSA}{MSW}$
Within groups	$n - c$	SSW	$MSW = \dfrac{SSW}{n-c}$	
Total	$n - 1$	SST		



- In Questions
  - You may have to create the ANOVA Table, or
  - You may be given an ANOVA Table to interpret

- There could be an additional column with p-value of the test statistic
- Remember: ANOVA uses one tailed F-Test (Entire α proportion is in the right tail)

ANOVA F-Test assumptions

- Independence
  - Samples are independent of each other. Measurement of one is not related to the measurement in the other
- Normality
  - Group populations are normally distributed
  - Test is relatively robust and works well for approximately normal populations
- Homogeneity of variance
  - Population variances are same
  - For same groups sizes, the test is quite robust against small variations
  - For different group sizes, this can bias our results
  - Test of equal variance is an important requirement
  - Levine test (F-test on absolute deviations from group medians can be used) to test of equal variances

When to use which test

- Compare two populations
  - Paired t-test for mean differences between two independent populations
  - Paired t-test for mean differences between two dependent populations
  - F-test for differences in variances between two independent populations
  - Z-test for differences between two proportions

- More than two populations
  - One-Way ANOVA for difference in independent population means

Multiple Mean Comparisons

- Finding out if the means are different is the first step
- If the null is rejected. The second step is to find out, which of the means are different.
- Instead of doing individual paired t-tests (for difference in means) for all the possible pair of means. We can run a joint multiple Tukey-Kramer Procedure.
- It has two steps
  - Compute the absolute mean differences $\left|\bar{X}_j - \bar{X}_i\right|$ $where\ i\ and\ j\ refer\ to\ differnt\ groups\ (i\ \neq j)$
  - How many pairs would have in c groups?
  - c(c-1)/2 (why?)
- Compute critical range for the Tukey Kramer procedure for each pair of sample sizes

CRITICAL RANGE FOR THE TUKEY-KRAMER PROCEDURE

$$\text{Critical range} = Q_\alpha \sqrt{\frac{MSW}{2}\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)} \qquad (10.13)$$

where

n_j = the sample size in group j
$n_{j'}$ = the sample size in group j'
Q_α = the upper-tail critical value from a **Studentized range distribution** having c degrees of freedom in the numerator and $n - c$ degrees of freedom in the denominator.

- Declare a group pair to be different if the absolute mean difference is more than the critical range

Multiple Mean Comparisons – Stores example

- Q_α is 4.05
- Pair-wise ranges

1. $|\bar{X}_1 - \bar{X}_2| = |29.982 - 31.994| = 2.012$
2. $|\bar{X}_1 - \bar{X}_3| = |29.982 - 30.912| = 0.930$
3. $|\bar{X}_1 - \bar{X}_4| = |29.982 - 30.334| = 0.352$
4. $|\bar{X}_2 - \bar{X}_3| = |31.994 - 30.912| = 1.082$
5. $|\bar{X}_2 - \bar{X}_4| = |31.994 - 30.334| = 1.660$
6. $|\bar{X}_3 - \bar{X}_4| = |30.912 - 30.334| = 0.578$

$$\text{Critical range} = 4.05\sqrt{\left(\frac{0.0439}{2}\right)\left(\frac{1}{5} + \frac{1}{5}\right)} = 0.3795$$

- Which of the groups are different?
- Only in #3 pair, there is no difference

Cumulative Probabilities = 0.95
Upper-Tail Area = 0.05
Numerator df_1

Denominator df_2	2	3	$\boxed{4}$	5	6	7	8	9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13
15	3.01	3.67	4.08	4.37	4.60	4.78	4.94	5.08
$\boxed{16}$	3.00	3.65	$\boxed{4.05}$	4.33	4.56	4.74	4.90	5.03

Studentized range for 0.05 significance, with 4 and 16 Numerator and denominator degrees of freedom

Q&A