13 Dec.

INFERENTIAL STATISTICS

* Confidence Interval:

Qn. Find the sample mean, variance & std. for the following dataset: {25, 26, 32, 37, 40}.

→ $\bar{x} = \frac{25+26+32+37+40}{5} = 32$

$Var = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{(25-32)^2+(26-32)^2+(32-32)^2+(37-32)^2+(40-32)^2}{4}$

$= 43.5$

$std.dev = \sqrt{43.5} = 6.595$

Qn. Among various ethnic groups, the SD of heights is known to be approximately 3 inches. We wish to construct a 95% confidence interval for the mean height of male Sweds(?). 48 male Sweds are surveyed. The sample mean is 71 inches.

$\bar{x} = 71$; $\sigma = 3$; $n = 48$.

Confidence interval = 95%.

$CI = \bar{x} \pm Z(\alpha/2)\frac{SD}{\sqrt{n}}$ $\alpha = 100-95 = 0.05$.

$= 71 \pm Z(0.025) \times \frac{3}{\sqrt{48}}$ $Z(1-0.025)$

$= Z(0.975)$.

$= 71 \pm \frac{1.96 \times 3}{\sqrt{48}} = 71 \pm 0.849$ $= 1.9 + 0.06$

$= 1.96$

$= (71.849, 70.151)$.

_/_/__

* Estimation of Parameters:

1. Point estimation.
2. Interval estimation.

Estimator: Rule / Formula / Function that tells how to calculate an estimate.

Estimate: Estimated numerical value from the estimator.

1. Point estimation:

A single value which is calculated for the sample data as an estimate for the unknown population parameter.

2. Interval estimation:

Interval estimate has a high probability of containing the parameter.

· Confidence Interval:

estimate $\pm$ (Critical value) · (Std. dev).

Qu. Plasma Aldosterone in dogs: Aldo is a hormone. 8 dogs with heart failure were treated with drug Captopril, & plasma concentrations of aldo were measured before & after the treatment:

Dogs	1	2	3	4	5	6	7	8
Before	210	219	222	224	310	345	360	400
After	202	211	210	211	289	310	323	339.

Suppose before-after change has normal dist. with $\sigma = 15$.

__/__/__

1. Display all values on a single graph, with paired values connected by a line.
2. Find the 95% CI for the mean change. [-21.4 ± 12.9]

95% = 100-95% = 0.05

3 Jan.

Qn. The National Student Loan survey: related to amount of money ~~borrowed~~ that borrowers owe. Sample of 1280 borrowers.
Mean debt : 18,900 \$.
Standard deviation : 49,000 \$.
95% confidence interval : $\alpha = 0.05$.

$\bar{x} = 18\,900$; $\sigma = 49{,}000$; $\alpha = 0.05$; $n = 1280$.

$$CI = \bar{x} \pm z(\alpha/2) \frac{SD}{\sqrt{n}}$$

$$= 18900 \pm z(1-0.025) \frac{49{,}000}{\sqrt{1280}}$$

$$= 18900 \pm 1.96 \cdot \frac{49\,000}{35.777}$$

$$= 18\,900 \pm \frac{96\,040}{35.777} = 18900 \pm 2684.4.$$

$$= ~~(21584.4,~~ (16\,215.6, 21\,584.4).$$

If $n = 320$:

$$CI = 18900 \pm 1.96 \cdot \frac{49000}{\sqrt{320}}$$

$$CI = 18900 \pm 5368.8.$$

$$= (13\,531.2, 24\,268.8).$$

$\Rightarrow$ If n decreases, the interval range increases.

__/__/____

$n = 1280$; confidence level: 99%

$\alpha = 100 - 99\% = 0.01$

$\alpha/2 = 0.005$

$z(1-\alpha/2) = z(0.995) = 2.58$ or 2.57. ?

$CI = \bar{x} \pm z(1-\alpha/2) \cdot \frac{SD}{\sqrt{n}}$

$CI = 18900 \pm 2.58 \cdot \frac{49000}{\sqrt{1280}}$

$= 18900 \pm \frac{126420}{\sqrt{1280}}$

$= 18900 \pm 3533.55$

$= (15366.45, 22433.55)$.

When confidence level increases, the range increase!

Qn. There are 250 cats. Avg weight 12 kg. Std. dev = 8 kg. Sample 4, what is the probability they have an average weight of greater than 10 kg but less than 25 kg.

$\mu = 12$; $\sigma = 8$; $n = 4$; $\sigma_{\bar{x}} = 8/\sqrt{4} = 8/2 = 4$

$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$

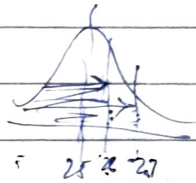
$z_{10} = \frac{10-12}{4} = \frac{-2}{4} = \frac{-1}{2}$; z value $= 0.3085$.

$z_{25} = \frac{25-12}{4} = \frac{13}{4} = 3.25$; z value $= 0.9994$.

∴ b/w 10-25 = $Z_{25} - Z_{10} = 0.9994 - 0.3085$
$= 0.6909$.

≈ 69.09%

Qn. Avg age : Women : 25.
Men : 27.8.
Std. dev : Women : 4.
n = 32.
b/w : 26 & 27.



$\mu = 25$; $\sigma = 4$; $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 4/\sqrt{32} = 0.707$.

$Z_{26} = \frac{26-25}{0.707} = \frac{1}{0.707} = 1.414$; Zvalue = 0.9207.

$Z_{27} = \frac{27-25}{0.707} = \frac{2}{0.707} = 2.83$; Zvalue = 0.9977.

$$\begin{array}{r} 0.9977 - \\ 0.9207 \\ \hline 0.0770 \end{array}$$

b/w : 26-27 = $Z_{27} - Z_{26} = 0.9977 - 0.9207$
$= 0.077$.

≈ 7.7%

__/__/__

10 Jan.

* Sample distribution ≠ Sampling distribution.

* Precision & Accuracy:

Precision: How close together repeated measurements or estimates are regardless of whether they are close to the true value.
- Reflects degree of variability in the sample data.

High precision: narrow Confidence Interval.
Low precision: Wide Confidence Interval.

Accuracy: Refers to how close a measurement or estimate is to the true value or the true population parameter.

Margin of error: $z * \frac{\sigma}{\sqrt{n}}$.

Sample size inversely proportional to the margin of error.

Qn. Create a CI to estimate μ driving range.
Margin of error to be no more than 10 km at 90% CI.
Std. dev: 15 km.

$z * \frac{15}{\sqrt{n}} \leq 10.$ $\alpha = \cancel{90\%}\ 100-90 = 0.1$

$z(1-\alpha/2) = z(1-0.05) = z(0.95) = 1.65$

$ME = z * \sigma/\sqrt{n}. \Rightarrow \sqrt{n} = \frac{z_{\alpha/2}\,\sigma}{ME}.$

$\therefore n = \left(\frac{z_{\alpha/2}\,\sigma}{ME}\right)^2 = \frac{1.65 \cdot 15}{10} = \left(\frac{24.75}{10}\right)^2 = \frac{612.5625}{100} = 6.12$

RANDOM

__/__/__

Qn: $\mu = 98.6°F$. Std dev $= 0.62°F$. $n = 106$.
Find probability of a mean of $98.2°F$ or lower.

$\frac{z\sigma}{\sqrt{n}}$.

$$z = \frac{\bar{x}-\mu}{\sigma} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}},$$

$$z = \frac{98.2-98.6}{0.62/\sqrt{106}} = \frac{-0.4}{0.06022} = \underline{-6.642}.$$

$\therefore$ z value ≈ 0.

Probability $= 0.001$

13 Jan.

* Hypothesis Testing:

Hypothesis: A claim / statement about a population parameter.

H_0: Null hypothesis: There is no significant difference.
H_1: Alternate hypothesis: Contradicts the null hypothesis.

Two tailed: $H_0: \mu = 23$.
$H_1: \mu \neq 23$.

Onetailed: Left-tailed: $H_0: \mu \geq 23$.
$H_1: \mu < 23$.

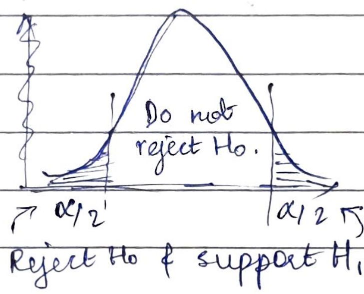
Right-tailed: $H_0: \mu \leq 23$.
$H_1: \mu > 23$.

Note: ~~Note~~ H_0 always has '=' sign.

If we reject H_0, we can support H_1.
If we fail to reject H_0, we cannot support H_1.

Significance level: α: Specifies the size of the rejection region.

95% confidence level ⇒ 5% significance = $\alpha = 5\%$.



Two tailed.

Right tailed: More, exceed, over, higher, above, beyond, larger, increased.

Qn. In recent years, the μ age of all clg students in city X has been 23.
$n = 42$ → $\mu = 23$. Std. dev. $\sigma = 2.4$; $\bar{x} = 23.8$
$\alpha = 0.05$. Normal distribution.
Can we infer that the population mean age has changed?

→ 2 tailed — ∵ ~~normal~~ 'age has changed'.
$\alpha = 0.05$; $\alpha/2 = 0.025$.

H_0 : $\mu = 23$.
H_1 : $\mu \neq 23$.

__/__/__

NOTE : z test used : σ known.
t test used : σ unknown.

According to the Central Limit Theorem, if sample is large ($n \geq 30$), use z-test.

Since $n \geq 30$, also σ is known, z-test.

$$\cancel{z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{23.8 - 23}{2.4/\sqrt{42}} = \frac{0.8}{0.37} = 2.16}$$

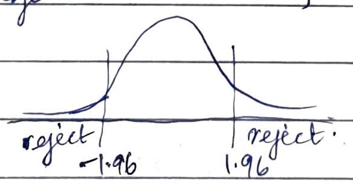
$\therefore$ z value = 0.9846.

~~z value = α z(1 − α/2) = z~~

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{23.8 - 23}{2.4/\sqrt{42}} = \frac{0.8}{0.37} = \underline{2.16}.$$

z : $z(1-\alpha/2) \Rightarrow$ ~~z(1-α/2)~~ = 0.945 = 1.96.

$\therefore$ Range : −1.96 & 1.96.



Since z = 2.16, 2.16 > 1.96, reject Ho.

$\therefore$ We can say the mean age has changed.

z(0.01) = −2.33.

When α = 0.02 ? $z(1-\alpha/2) = z(1-0.01) = z(0.99) = 2.33$

2.16 Range : −2.33 ~ 2.33.

Since ~~2.16~~ < 2.33, we cannot reject Ho.

At α = 0.02, there is ~~not~~ enough evidence that the mean age has changed.

__/__/__

14 jan.

Qu. A telecom service provider.

$\mu = 400$ rs/month. Std. dev. $\sigma = 25$ rs. $n = 50$.

$\bar{x} = 250$ $\sigma = 15$.

What to say with respect to the claim made by the service provider?

$H_0 : \mu = 400$.

$H_1 : \mu \neq 400$.

$\sigma = 25$.

$$z = \frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} = \frac{250-400}{25/\sqrt{50}} = \frac{-150}{3.54} = -42.34$$

z critical $\Rightarrow \alpha = 5\%$ $\rightarrow$ $\alpha = 0.05$. $\alpha/2 = 0.025$.

$\Rightarrow z(1-\alpha/2) = z(0.975) = 1.96$..

$\Rightarrow -1.96 < z < 1.96$..

Since $z = -42.34$ is not b/w -1.96 & 1.96, we cannot reject H_0.

16 Jan.

Qu. User exercise: $\mu \leq 30$ min/day..

$n = 50$. $\bar{x} = 32$ min/day. $\sigma = 6$ min.

$\alpha = 3\% = 0.03$. $1-\alpha = 0.97$.

$\mu \leq 30$ $\rightarrow$ right tailed.

$$z = \frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} = \frac{32-30}{6/\sqrt{50}} = \frac{2}{6/\sqrt{50}} = 2.36$$

z CV $\Rightarrow z(1-\alpha) = z(0.97) = 1.88$.

$\Rightarrow$ Since $2.36 > 1.88$, we reject $H_0 : \mu \leq 30$.

_/_/_

What happen when $\sigma \uparrow = 8$.

$$z = \frac{32-30}{8/\sqrt{50}} = \frac{1}{4/\sqrt{50}} = \frac{1}{0.5657} = 1.77$$

Since $1.77 < 1.88$, we ~~acce~~ ~~accept Ho.~~ cannot reject Ho. $\mu' \leq 30$.

17 Jan.

* Errors:

- Type I error: Rejecting Ho when it is actually True.
- Type 2 error: Not rejecting Ho when it is actually false.

P(type I error / Ho is true) = α.
P(type II error / Ho is false) = β.

P(rejecting a false Ho) = $1-\beta$.

	Ho True	Ho False
Reject Ho	Type I error	✓
Fail to reject Ho	✓	Type II error

→ Reducing type I error: reducing significance level / increasing confidence level.
↳ Prescriptive testing.

→ Reducing type 2 error: Descriptive testing.

- Type III error: When Ho is rejected for the wrong reason.

* t - distribution:

In z-test, we use population's standard deviation. $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

In t-test, we use sample's standard deviation.

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}.$$

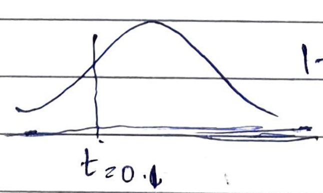
The degree of freedom = $n-1$.

t-distribution used when sample size is low & population's std. dev. is unknown.

[∵ Central Limit Theorem: when sample size is large, it is close to normal distribution, where z-test can be used].

→ $n = 15$; Conf. lvl = 90% — left tailed.

df = 14.



$1 - 0.9 = 0.1$

$t = -1.345$.

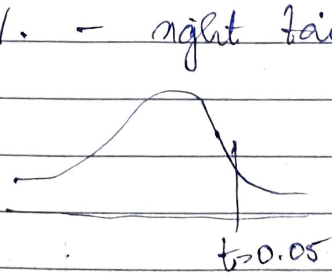
[-ve because t-value < 0.5].

→ $n = 25$, Conf lvl: 95% — right tailed.

df = 24.

$1 - 0.95 = 0.05$.

$t = 1.711$.



→ $n = 35$. Conf lvl: 99% – two tailed.

$df = 34$.

$1 - 0.99/2 = 0.01/2 = 0.005$.

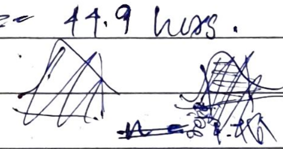
→ $1 - 0.005 = \underline{0.995}$

→ tvalue $= \underline{2.728}$

Qn. One Sample t-test:

Company wants to test the claim that their battery lasts more than 40 hours.

$n = 15$; $\bar{x} = 44.9$ hrs. $s = 8.9$ hrs.

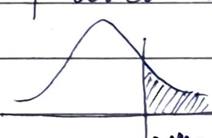


Conf level: 95%.

$\mu > 40$.

~~Rightleft~~ Right tailed.

Right tailed



$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{44.9 - 40}{8.9/\sqrt{15}} = \frac{4.9}{2.3} = \underline{2.13}$$

$\alpha = 1 - 0.95 = 0.05$

→ $t_{cv} = 1.761$.

Since tvalue > cv $(2.13 > 1.761)$, we ~~cannot~~ ~~cannot~~ reject H_0.

$\because H_1 = \mu > 40$; ⇒ $H_0 = \mu \le 40$. → Right tailed.

/ /

20 Jan.

~~t t~~

Qn Random sample $n = 21$. from a large population $\mu = 22$. $\sigma = 4.8$. $\alpha = 0.01$.
Can we conclude that the population mean is significantly below 24?

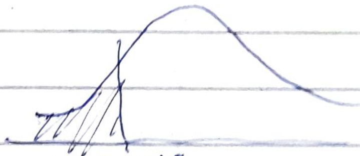
$n = 21$ $\quad \mu = 22$ $\quad s = 4.8$ $\quad \alpha = 0.01$.

$H_1 : \mu < 24$.

$H_0 : \mu \geq 24$. Left-tailed.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{22-24}{4.8/\sqrt{27}} = \frac{-2}{0.924} = -2.1645$$

$t\ CV = -2.479$.



-2.479

Since $-2.1645 > -2.479$, we cannot reject H_0.

__/__/__

21 Jan.

* Choosing t-test:

If samples sets are same or related
↳ Paired (dependent) t-test.

If no:
samples are of same size:
↳ equal variance (independent).

If no:
sample variance are equal??:
↳ equal variance (independent).

If not, unequal variance (independent).

Qn. Paired t-test:
~~Rating magazines~~
Exercise equipments, measuring heart~~bea~~ rate (bpm), equipments A & B compare the mean of the difference b/w the pair of sample data (is it 0?).

Person	1	2	3	4	5
A	61	72	66	88	75
B	55	75	68	81	78
$D_i = A_i - B_i$	6	-3	-2	7	-3

→ $t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{\Sigma d_i/n}{S_d/\sqrt{n}} = \frac{\Sigma(x_1 - x_2)/n}{S_d/\sqrt{n}}$.

$t = \frac{5/5}{S_d/\sqrt{n}}$ $H_0: \mu_D = 0$. $H_1: \mu_D \neq 0$.

$$s_d = \sqrt{\frac{\Sigma(x_i - \mu)^2}{n-1}} = \sqrt{\frac{25+16+9+36+16}{4}}$$

$$= \sqrt{\frac{102}{4}} = 5.05$$

$$t = \frac{1}{5.05/\sqrt{5}} = 0.443$$

df 4; taking $\alpha = 0.5 \Rightarrow \alpha/2 = 0.025$.

CV = 2.776.

Range: $-2.776 < 0.443 < 2.776$.

Since $t <$ CV ($0.443 < 2.776$), cannot reject H_0.

Qn. Consumer rating magazine.
Compare mean life length of 2 batteries.
$n = 5$.

Toy	1	2	3	4	5	
Battery 1.	52.6	103.4	68.2	88.4	111.6	hrs.
Battery 2.	61.4	112.8	67.1	92.3	121.5	hrs.

a. State H_0 & H_1.
b. Identify test statistic & critical region ($\alpha = 0.05$).
c. State conclusion, find p-value.
d. Construct a CI that is relevant to the problem; use it to verify the ~~ob~~ conclusion.

→ $H_0 : \mu_D = 0$; $H_1 : \mu_D \neq 0$.

	1	2	3	4	5
d_i =	-8.8	-9.4	1.1	-3.9	-9.9

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-6.18}{s_d/\sqrt{n}}.$$

_/_/_

$$Sd = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{89.268}{4}} = 4.72$$

$$t = \frac{-6.18}{4.72/\sqrt{5}} = -2.93$$

df = 4 ; $\alpha = 0.05$; $\alpha/2 = 0.025$.

CV = 2.776.

Range = $-2.776 \sim 2.776$.

Since $-2.93 < -2.776$; we reject H_0.

P-value skip! ☺

Confidence Interval = $\bar{d} \pm t_{cv} \times \frac{Sd}{\sqrt{n}}$.

$$= -6.18 \pm -2.776 \cdot \frac{4.72}{\sqrt{5}}$$

$$= -6.18 \pm -5.86.$$

$$= -12.04 \sim -0.32.$$

DA fam:

P-value: using df & t value.

4   2.93.

In the t table, find t value 2.93.

$\alpha/2 \rightarrow$	0.025	0.02	0.01
df = 4 $\Rightarrow$	2.776	2.93	3.747.

So, we take the $\alpha/2$ value close to 2.776,

$\therefore$ p value$_{/2}$ = 0.02.

$\therefore$ p value = $0.02 \times 2 = 0.04$.

Since $0.04 < 0.05$; we reject H_0.

p value < α

/ /

* Equal variance: Independent samples t-test.

Qu. A statistics teacher wants to compare his 2 classes to see if they performed any differently on the test he gave that semester. Class A had 25 students with an average score of 70, $\sigma = 15$; Class B had 20 students with $\mu_B = 74$, $\sigma = 25$. Using $\alpha = 0.05$, did the 2 classes perform differently on the tests?

→ $\mu_A = 70$; $\mu_B = 74$. $\alpha = 0.05$.

$\sigma_A = 15$; $\sigma_B = 25$.

$n_A = 25$; $n_B = 20$.

$H_0: \mu_A - \mu_B = 0.$

$H_1: \mu_A \neq \mu_B$ or $\mu_A - \mu_B \neq 0.$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \quad ; \quad S_p^2 = \frac{S_1^2 df_1 + S_2^2 df_2}{df_1 + df_2}.$$

$$S_p^2 = \frac{15^2 \cdot 24 + 25^2 \cdot 19}{24+19} = \cancel{19.42} \quad \underline{401.744}.$$

$$\cancel{t = \frac{70-74}{\sqrt{\frac{19.42}{25} + \frac{19.42}{20}}} = \frac{-4}{1.322} = -3.025}$$

$$t = \frac{70-74}{\sqrt{\frac{401.744}{25} + \frac{401.744}{25}}} = \underline{-0.66522}.$$

$\alpha/2 = 0.025$. $df_{total} = 43$. ⇒ $t_{cv} = \underline{2.0167}$

Since $-2.0167 < t_{value} < 2.0164$ not true; $-0.66522 < -2.0167$, reject H_0.

__/__/__

Qn. Unequal variance: Prev question:

$\mu_A = 70$ $\quad$ $\mu_B = 74$. $\quad$ $\alpha = 0.05$.

$\sigma_A = 15$ $\quad$ $\sigma_B = 25$.

$n_A = 25$ $\quad$ $n_B = 20$. $\quad$ $H_0: \mu_1 - \mu_2 = 0$.

$df_1 = 24$. $\quad$ $df_2 = 19$. $\quad$ $H_1: \mu_1 - \mu_2 \neq 0$.

* ANOVA:

Qn.

Sno.	Method A	Method B	Method C
1	10	8	9
2	9	9	8
3	8	10	7
4	7.5	8	10
5	8.5	8.5	9
6	9	7	8
7	10	9.5	7
8	8	9	10
9	8	7	9
10	9	10	8
Group Mean:	8.7	8.6	8.5

Overall Mean: 8.6

$SB = n\left[\Sigma(\mu_c - \mu_g)^2\right] = 10\left[(8.7-8.6)^2 + (8.6-8.6)^2 + (8.5-8.6)^2\right]$

$= 10[0.01 + 0.01] = 10 \times 0.02 = \underline{0.2}$

$SW = \Sigma\Sigma(x_{ij} - x_c)^2$.

$SW_A = [(10-8.7)^2 + (9-8.7)^2 + (8-8.7)^2 + (7.5-8.7)^2 + (8.5-8.7)^2 + (9-8.7)^2 + (10-8.7)^2 + (8-8.7)^2$
$+ (8-8.7)^2 + (9-8.7)^2] = \underline{6.6}$.

$SW_B = [2(8-8.6)^2 + 2(9-8.6)^2 + 2(10-8.6)^2 + (8.5-8.6)^2 + 2(7-8.6)^2$
$+ (9.5-8.6)^2] = \underline{10.9}$.

$SW_C = [3(9-8.5)^2 + 3(8-8.5)^2 + 2(7-8.5)^2 + 2(10-8.5)^2]$
$= \underline{10.5}$.

$\therefore SW = 10.5 + 10.9 + 8.6 = \underline{28}$.

$ANOVA = \frac{SB}{SW} = \frac{0.2}{28} = 0.00714285714\dot{3}$.

Since statistic value < 1, we **fail to reject H_0**.

Assuming $\alpha = 0.05$;
$F_{stat} = 0.0071$.
F_{cv} ~~table~~ $\left(\frac{df_1}{df_2}\right) = \frac{\text{no. of } \text{classes} \text{ ~~samples~~} - 1}{\text{no. of elements} - \text{no. of classes}}$.

~~$F_{cv} = \frac{3-1}{30-3} = \frac{2}{27} = 0.074$~~ $F(2, 27)$

$F_{cv} = F(2,27) = 3.35$.

28 Jan.

SSB = Variance b/w the groups.
SSW = Variance within groups.

$MSB = \frac{SSB}{df_B}$ $MSW = \frac{SSW}{df_w}$.

$F \text{ statistic} = \frac{MSB}{MSW}$.

28 Jan

Qn. Researchers want to test a new anti-anxiety medication. They split participants into 3 conditions (0mg, 50mg, and 100mg), then ask them to rate their anxiety level on a scale of 1-10. Are there any differences b/w the 3 conditions, using $\alpha = 0.05$?

0mg	50 mg	100 mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

→ $\mu_1 : 8.14$ $\qquad \mu_2 : 6.71$ $\qquad \mu_3 = 3$.

① Since levels are independent, One-way ANOVA.

② ~~Overall Mean~~ $H_0 : \mu_1 = \mu_2 = \mu_3$.

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$.

③ $\mu_1 = 8.14$; $\mu_2 = 6.71$; $\mu_3 = 3$; Overall $\mu_g = 5.95$.

④ $SSB = n\sum(\bar{x}_i - \bar{x})^2 = 98.5334$

$SSW = \sum\sum(x_{ij} - \bar{x}_i)^2$

$SSW_1 = 2.8572$

$SSW_2 = 3.4287$

$SSW_3 = 4$

$\Big\}\ SSW = 10.2859$

_/_/_

(5) $df_B = k-1 = 3-1 = 2$.
df_W = total elements − k = 21 − 3 = 18.

(6) $MSB = \frac{SSB}{df_B} = \frac{98.5334}{2} = 49.2667$.

$MSW = \frac{SSW}{df_W} = \frac{10.2859}{18} = 0.5714$.

(7) $F_{statistic} = \frac{MSB}{MSW} = \frac{49.2667}{0.5714} = 86.221$.

(8) $F_{CV}(2,18) = 3.55$.

(9) If $F_{cv} > F_{stat}$, fail to reject H_0.
If $F_{cv} < F_{stat}$, reject H_0.

⇒ Reject H_0.

30 Jan.

* ~~One~~ 2-Way ANOVA:

~~Drug A~~

	Drug A	Drug B	
Male	6	4	
	4	5	
	7	6	
	9	4	
	3	5	
Mean:	5.8	5.4	$M_m = 5.6$.
Female	8	3	
	3	5	
	5	9	
	8	2	
	6	3	
Mean:	6	4.4	$M_f = 5.2$.

__/__/__

Drug A : Male : 5.8 Drug B: Male : 5.4.
Female : 6. Female : 4.4.
Total : 5.9. Total : 4.9.

Male total = 5.6 ; Female total = 5.2.
Total = $\underline{5.4}$.

$SS_{b/w} = n \sum\sum (\mu_{gi} - \mu_{total})^2$
$= 5 \left[(5.8-5.4)^2 + (5.4-5.4)^2 + (6-5.4)^2 + (4.4-5.4)^2 \right]$
$= 5 \left[0.16 + 0 + 0.36 + 1 \right] = \underline{7.6}$.

$df_{b/w} = p \cdot q - 1 = 2\times 2 - 1 = 4 - 1 = 3$.

$\sigma_{b/w} = \frac{7.6}{3} = \underline{2.53}$.

p = factoring acc to drugs. ; q = factoring acc to gender.

$SS_A = n \cdot q \sum (\mu_A - \mu_{total})^2$ (→ drugs.)
$= 5 \cdot 2 \left[(5.9-5.4)^2 + (4.9-5.4)^2 \right]$.
$= \underline{5}$.

$df_A = p - 1 = 2 - 1 = \underline{1}$.

$\therefore \sigma_A = \frac{5}{1} = 5$.

$SS_B = n \cdot p \sum (\mu_B - \mu_g)^2$. (→ male & female.)
$= 5 \cdot 2 \left[(5.6-5.4)^2 + (5.2-5.4)^2 \right]$
$= \underline{0.8}$

$df_B = 2 - 1 = 1$.

$\therefore \sigma_B^2 = \underline{0.8}$.

3 Jan.

$SS_{tot} = SS_A + SS_B + SS_{AB} + SS_{err}$

SS_A ↓ Variance that can be explained by factor A.

SS_B ↓ By factor B.

SS_{AB} ↓ By interaction of A & B.

SS_{err} ↳ error variance.

$$SS_{tot} = \underbrace{SS_A + SS_B + SS_{AB}}_{\text{Between-class}} + \underbrace{SS_{err}}_{\text{within}}$$

$$SS_{AB} = SS_{btw} - SS_A - SS_B = 7.6 - 5 - 0.8 = \underline{1.8}$$

$$df_{AB} = df_A \times df_B = (p-1) \times (q-1) = \underline{1}$$

$$\sigma^2_{AB}\ (MSB_{AB}) = \frac{SS_{AB}}{df_{AB}} = \frac{1.8}{1} = \underline{\underline{1.8}}$$

$$SS_{err} = \sum\sum\sum (x_{mij} - \overline{AB}_{ij})^2$$

$$= \sum\sum\sum (\text{element} - \mu_{\text{classes}\,j})^2$$

$$= \Big[(6-5.8)^2 + (4-5.8)^2 + (7-5.8)^2 + (9-5.8)^2 + (3-5.8)^2 +$$
$$(4-5.4)^2 + (5-5.4)^2 + (6-5.4)^2 + (7-5.4)^2 + (5-5.4)^2 +$$
$$(8-6)^2 + (3-6)^2 + (5-6)^2 + (8-6)^2 + (6-6)^2 +$$
$$(3-4.4)^2 + (5-4.4)^2 + (9-4.4)^2 + (2-4.4)^2 + (3-4.4)^2 \Big]$$

$$= \Big[0.04 + 3.24 + 1.44 + 10.24 + 7.84 +$$
$$1.96 + 0.16 + 0.36 + 2.56 + 0.16 +$$
$$4 + 9 + 1 + 4 + 0 +$$
$$1.96 + 0.36 + 21.16 + 5.76 + 1.96 \Big]$$

$$= \underline{\underline{77.2}}$$

_/_/__

$df_{err} = (n-1)\,p\cdot q = (5-1)\,2\cdot 2 = 4\cdot 2\cdot 2 = \underline{\underline{16}}.$

σ^2_{err} [or MS_W or MS_{err}]

$$\sigma^2_{err} = \frac{SS_{err}}{df_{err}} = \frac{77.2}{16} = \underline{\underline{4.825}}.$$

$$F_A = \frac{MSB_A\,(\sigma^2_A)}{MSW_A\,(\sigma^2_{err})} = \frac{5}{4.825} = \underline{\underline{1.04}}.$$

$\alpha = 0.05$:

$F_{critical} = F(df_A, df_{err}) = F(1,16) = \underline{\underline{4.49}}.$

Since $F_{critical} > F_A$, we fail to reject H_0.

$$F_B = \frac{MSB_B\,(\sigma^2_B)}{MSW_B\,(\sigma^2_{err})} = \frac{0.8}{4.825} = 0.166$$

$F_{critical} = F(df_B, df_{err}) = F(1,16) = \underline{\underline{4.49}}.$

Since $F_{critical} > F_B$, we fail to reject H_0.

$$F_{AB} = \frac{MSB_{AB}\,(\sigma^2_{AB})}{MSW_{AB}\,(\sigma^2_{err})} = \frac{1.8}{4.825} = \underline{\underline{0.37}}.$$

$F_{critical} = F(df_{AB}, df_{err}) = F(1,16) = \underline{\underline{4.49}}$

Since $F_{critical} > F_{AB}$, we fail to reject H_0.

Note: [Note: $SS_{tot} = \sum\sum\sum(\text{element} - \mu_{global})^2.$

$df_{tot} = (n\cdot p\cdot q) - 1$

$\sigma^2_{tot} = \dfrac{SS_{tot}}{df_{Tot}}$.]

4 Feb.

__/__/__

Qn: Does Noise have an effect on the marks ~~of~~ a student scores?

Students	Low noise	Medium Noise	High noise.
Male	10	7	4
	12	9	5
$\mu_M = 8.17$.	11	8	6
	9	12	5
	$\mu_{lm} = 10.5$	$\mu_{mm} = 9$.	$\mu_{hm} = 5$.
Female	12	13	6
	13	15	6
$\mu_F = 10$.	10	12	4
	13	12	4
	$\mu_{lf} = 12$.	$\mu_{mf} = 13$.	$\mu_{hf} = 5$.
	$\mu_{low} = 11.25$.	$\mu_{med} = 11$.	$\mu_{high} = 5$.

→ Anova [∵ ≥ 2 class].
2way [∵ > 1 factor].

Overall $\mu_g = 9.1$.

Hypothesis: H_0: No significant difference b/w the ~~H~~ groups of noise.
H_1: There is ~~some~~ significant difference b/w ~~the atleast 2 groups.~~ the groups of noise.

Male: Low: 10.5
Med: 9
High: 5.

Female: Low: 12
Med: 13
High: 5.

Total: Low = 11.25 Med = 11 High = 5.
Male = 8.17 Female = 10 Overall = 9.1.

$$SS_{tot} = SS_{B/W} + SS_{within}.$$
$$(SS_{B_A} + SS_{B_B} + SS_{B_{AB}})$$

_/_/_

EXTRA, NOT NEEDED

$$SS_{b/w} = n \sum\sum (\mu_{class} - \mu_{global})^2$$
$$= 4[(10.5-9.1)^2 + (9-9.1)^2 + (5-9.81)^2 + (12-9.1)^2 + (13-9.1)^2 + (15-9.1)^2]$$
$$= 4[1.96 + 0.01 + 16.81 + 8.41 + 15.21 + 1\cancel{6}0.81]$$
$$= \cancel{4 \times 77.21 = 308.84}.\ 4 \times 59.21 = 236.84.$$

$$df_{b/w} = (p \cdot q) - 1 = (2 \cdot 3) - 1 = 6 - 1 = 5$$

$$\sigma^2_{b/w} = \frac{SS_{b/w}}{df_{b/w}} = \frac{\cancel{308.84}}{5} = \cancel{61.768}.\ 47.368.$$

~~p-factors q-classes~~. p-classes; q-factors (gender).

$$SS_{B_A} = nq \sum(\mu_{classes} - \mu_{total}).$$
$$= 4 \cdot 2 \,[(11.25 - 9.1)^2 + (11-9.1)^2 + (5-9.1)^2].$$
$$= 8[4.6225 + 3.61 + 16.81]$$
$$= 200.34.$$

$$df_A = p - 1 = 2.$$

$$MS_A = \sigma^2 = \frac{SS_{B_A}}{df_A} = \frac{200.34}{2} = 100.17.$$

$$SS_{within} = SS_{err} = \sum\sum\sum(\cancel{x_{class}}\ element - \mu_{class_i})^2.$$
$$= [(10-10.5)^2 + (12-10.5)^2 + (11-10.5)^2 + (9-10.5)^2 +$$
$$(7-9)^2 + (9-9)^2 + (8-9)^2 + (12-9)^2 +$$
$$(4-5)^2 + (5-5)^2 + (6-5)^2 + (5-5)^2 +$$
$$(12-12)^2 + (13-12)^2 + (10-12)^2 + (13-12)^2 +$$
$$(13-13)^2 + (15-13)^2 + (12-13)^2 + (12-13)^2 +$$
$$(6-5)^2 + (6-5)^2 + (4-5)^2 + (4-5)^2].$$
$$= [5 + 14 + 2 + 6 + 6 + 4]$$
$$= 37.$$

~~$df_{Brr} = q \cdot \ldots = 2 \ldots ; \ldots MS_{Brr} = \ldots 37$~~.

_/_/_

$df_{err} = (n-1)pq = (4-1)(3 \cdot 2) = 3 \cdot 3 \cdot 2 = 18$.

$\sigma^2_{err} = MS_{w} = MS_{err} = \frac{SS_{err}}{df_{err}} = \frac{37}{18} = 2.056$.

$F_{stat_A} = \frac{MS_A}{MS_{err}} = \frac{100.17}{2.056} = 48.721$.

At $\alpha = 0.05$:

$F_{CV} = F(df_A, df_{err}) = F(2, 18) = 3.55$.

Since $F_{CV} < F_{stat_A}$, we reject H_0. So, yes.

ii. Does gender affect ~~of~~ the student score?

$SS_B = np \sum (\mu_B - \mu_{global})^2$

$= 4 \cdot 3 [(8.17 - 9.1)^2 + (10 - 9.1)^2]$

$= 12(0.8649 + 0.81) = 1.6749 \times 12 = 20.0988$.

$df_B = q - 1 = 2 - 1 = 1$.

$\sigma^2_B = MS_B = \frac{SS_B}{df_B} = \frac{20.0988}{1} = 20.0988$.

$F_{stat_B} = \frac{MS_B}{MS_{err}} = \frac{20.0988}{2.056} = 9.78$.

~~Since FC~~

$F_{CV} = F(df_B, df_{err}) = F(1, 18) = 4.41$.

Since $F_{CV} < F_{stat_B}$, we reject H_0. So, yes.

$\therefore$ Both noise & gender groups do have significant differences.
And both noise & gender do affect ~~the~~ student's scores.