1. Show that when calculating a confidence interval for an unknown population proportion p, $E(\hat{p}) = p$ and $V(\hat{p}) = \frac{p(1-p)}{n}$.

$X = \#\text{"S"}$ in n trails.

$$\hat{p} = \frac{X}{n} \qquad E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} \cdot E(X) = \frac{1}{n} \times np = p$$

$$V(\hat{p}) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot V(X) = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n}$$

2. Of 48 engineers that were interviewed about the topic, 36 think that the new reinforcement of a concrete building is not earthquake proof. Give a 95% confidence interval for the true proportion of engineers that thinks that the new reinforcement is not earthquake proof.

p = true proportion of engineers ...... not proof

$$\hat{p} = \frac{36}{48} = 0.75$$

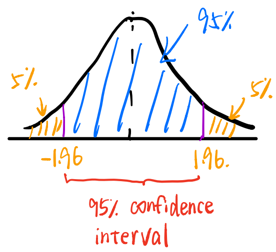
$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Use Formula: $\left[\hat{p} - \overset{-1.96}{z} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{\underset{48}{n}}}, \ \hat{p} + \overset{1.96}{z} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{\underset{48}{n}}}\right] = [0.6275, 0.8725]$

Why? ⟶ Start from Sampling Distribution

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Z:



$$-1.96 \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1.96$$

Use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ to approximate $\sqrt{\frac{p(1-p)}{n}}$

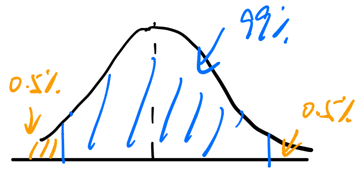
$$-1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \hat{p} - p \leq 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$-1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p - \hat{p} \leq 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\hat{p} - 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

3. Is it good to change speed limits at certain segments of streets? When a speed limit at a street segment was 50 km/h, the speeds of 100 randomly selected vehicles traversing the street was monitored and 49 violations of the speed limit were observed. After the speed limit was raised to 60 km/h, the speeds of another 100 randomly selected vehicles was monitored. Now only 19 violations of the speed limit were observed. Let p_1 and p_2 be the true proportions of speed limit violations before and after the speed limit change. Give a 99% confidence interval for $p_1 - p_2$.

Recall : $\left[(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]$



$\hat{P}_1 = \frac{49}{100} = 0.49 \quad \hat{P}_2 = \frac{19}{100} = 0.19$

$Z_{0.005} = -2.575 \quad Z_{1-0.005} = 2.575 \quad n_1 = n_2 = 100$

99% CI for P : $[0.136, 0.464]$

Detailed : $P\left(P_1 - P_2 \leq (\hat{P}_1 - \hat{P}_2) - Z\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}\right)$

$= P\left(P_1 - P_2 \leq 0.30 - 2.58\sqrt{\frac{0.49 \cdot 0.51}{100} + \frac{0.19 \times 0.81}{100}}\right)$

$= P(P_1 - P_2 \leq 0.136) = 0.025$

$P\left(P_1 - P_2 \leq (\hat{P}_1 - \hat{P}_2) + Z\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}\right)$

$= P\left(P_1 - P_2 \leq 0.30 + 2.58\sqrt{\frac{0.49 \cdot 0.51}{100} + \frac{0.19 \times 0.81}{100}}\right)$

$= P(P_1 - P_2 \leq 0.464) = 0.975$.

Why ? $\hat{P}_1 \sim N\left(P_1, \sqrt{\frac{P_1(1-P_1)}{n_1}}\right) \quad \hat{P}_2 \sim N\left(P_2, \sqrt{\frac{P_2(1-P_2)}{n_2}}\right)$

$\hat{P}_1 - \hat{P}_2 \sim N\left(P_1 - P_2, \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}\right)$

$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \qquad \underset{(Z_{0.005})}{-2.575} \leq Z \leq \underset{(-Z_{0.005})}{2.575}$

$\Rightarrow \quad -2.575 \leq \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \leq 2.575$

$(\hat{P}_1 - \hat{P}_2) - 2.575\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}} \leq (P_1 - P_2) \leq (\hat{P}_1 - \hat{P}_2) + 2.575\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$

4. A company manufactures fan blades on five production lines A, B, C, D and E. These lines produce fan blades at the same rate and volume. In a sample of $n = 103$ randomly selected *defective* fan blades, 15 were manufactured on line A, 27 on line B, 31 on line C, 19 on line D and 11 on line E. Let p_i be the true proportion of defective fan blades that is manufactured on line i, $i = A, B, C, D, E$.

(a) Find a 95% confidence interval for p_A.

(b) Find a 95% confidence interval for $p_A - p_B$.

(a) $\hat{P}_A = \frac{15}{103} \approx 0.146$. $Z_{\frac{5\%}{2}} = -1.96$. $Z_{1-\frac{5\%}{2}} = 1.96$

95% C.I. for P_A : $\left[\hat{P}_A - Z_{1-\frac{5\%}{2}} \cdot \sqrt{\frac{\hat{P}_A(1-\hat{P}_A)}{n}}, \hat{P}_A + Z_{1-\frac{5\%}{2}} \cdot \sqrt{\frac{\hat{P}_A(1-\hat{P}_A)}{n}}\right]$

$\Rightarrow [0.078, 0.214]$

(b) $\hat{P}_B = \frac{27}{103} \approx 0.262$ $Z_{\frac{5\%}{2}} = -1.96$. $Z_{1-\frac{5\%}{2}} = 1.96$

95% C.I. for $P_A - P_B$: $\left[(\hat{P}_A - \hat{P}_B) - Z_{1-\frac{5\%}{2}} \cdot \sqrt{\frac{\hat{P}_A(1-\hat{P}_A)}{n} + \frac{\hat{P}_B(1-\hat{P}_B)}{n}}, (\hat{P}_A - \hat{P}_B) + Z_{1-\frac{5\%}{2}} \cdot \sqrt{\frac{\hat{P}_A(1-\hat{P}_A)}{n} + \frac{\hat{P}_B(1-\hat{P}_B)}{n}}\right]$

$\Rightarrow [-0.237, 0.005]$

5. Do well-rounded people get fewer colds? A study conducted by scientists from three United States universities found that people who have only a few social outlets (at most 3) get more colds than those who are involved in a variety of social activities (at least 6). From a sample of 276 healthy people, $n_1 = 96$ had only a few social outlets and $n_2 = 105$ were busy with six or more activities. When these people were exposed to a cold virus, the following results were observed:

	Few social outlets	Many social outlets
Sample size	96	105
Percent with colds	62%	35%

(a) Construct a 99% confidence interval for the difference in the two population proportions.

(b) Does there appear to be a difference in the population proportions for the two groups?

(c) You might think that coming into contact with more people would lead to more colds, but the data show the opposite effect. How can you explain this unexpected finding?

(a) $Z_{\frac{1\%}{2}} = -2.575$ $Z_{1-\frac{1\%}{2}} = 2.575$

P_1 = true proportions with colds who have few social outlets

P_2 = true proportions with colds who have more social outlets

$\hat{P}_1 = 0.62$, $\hat{P}_2 = 0.35$, $n_1 = 96$, $n_2 = 105$

95% C.I. for $P_1 - P_2$:

$\left[(\hat{P}_1 - \hat{P}_2) - Z_{1-\frac{1\%}{2}} \cdot \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}, (\hat{P}_A - \hat{P}_B) + Z_{1-\frac{1\%}{2}} \cdot \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}\right]$

$\Rightarrow [0.095, 0.445]$

(b) Yes. Because 0 is not included in the interval.

(c)

6. Because of erratic rainfall, supplemental irrigation is required for producing several crops. A research team has developed five alternative water-management strategies for irrigating cropland. A random sample of 100 agricultural engineers was asked which of the strategies they believed would yield maximum productivity. They responded as follows:

Strategy	A	B	C	D	E
Frequency	17	27	22	15	19

(a) Find a 90% confidence interval for the true proportion of agricultural engineers who recommend strategy C.

(b) Find a 90% confidence interval for the difference between the true proportions of agricultural engineers who recommend strategies E and B.

(c) Find a 90% confidence interval for the difference between the true proportions of agricultural engineers who recommend strategies A and D.

(a) $\hat{p}_c = 0.22$ $\quad Z_{\frac{10\%}{2}} = -1.645 \quad Z_{1-\frac{10\%}{2}} = 1.645$, $n = 100$

Recall: $\left[\hat{p}_c - Z_{1-\frac{10\%}{2}} \cdot \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n}}, \ \hat{p}_c + Z_{1-\frac{10\%}{2}} \cdot \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n}} \right]$

90% C.I. for P_c : $[0.152, 0.288]$

(b) $\hat{p}_B = 0.27 \quad \hat{p}_E = 0.19$

Recall : 90% C.I. for $P_B - P_c$

$$\left[(\hat{p}_B - \hat{p}_c) - Z_{1-\frac{10\%}{2}} \cdot \sqrt{\frac{\hat{p}_B(1-\hat{p}_B)}{n} + \frac{\hat{p}_E(1-\hat{p}_E)}{n}}, \ (\hat{p}_B - \hat{p}_c) + Z_{1-\frac{10\%}{2}} \cdot \sqrt{\frac{\hat{p}_B(1-\hat{p}_B)}{n} + \frac{\hat{p}_E(1-\hat{p}_E)}{n}} \right]$$

$\Rightarrow [-0.031, 0.191]$

(c) $\hat{p}_A = 0.17$, $\hat{p}_D = 0.15$

Recall : 90% C.I. for $P_A - P_D$

$$\left[(\hat{P}_A-\hat{P}_D) - Z_{1-\frac{10\%}{2}} \cdot \sqrt{\frac{\hat{P}_A(1-\hat{P}_D)}{n} + \frac{\hat{P}_A(1-\hat{P}_D)}{n}},\ (\hat{P}_A-\hat{P}_D) + Z_{1-\frac{10\%}{2}} \cdot \sqrt{\frac{\hat{P}_A(1-\hat{P}_D)}{n} + \frac{\hat{P}_A(1-\hat{P}_D)}{n}}\right]$$

$$\Rightarrow [-0.073, 0.113]$$

7. (See the previous tutorial) A taxi company is trying to decide whether to purchase brand A or brand B tires for its fleet of taxis. To estimate the difference in the two brands, an experiment is conducted using 10 of each brand. The tires are run until they wear out. The results are:

Brand A: $\overline{x}_A = 36300$ kilometers, $s_A = 5000$ kilometers.
Brand B: $\overline{x}_B = 38100$ kilometers, $s_B = 6100$ kilometers.

Assuming that the populations to be approximately normally distributed, compute a 90% confidence interval for $\frac{\sigma_A}{\sigma_B}$.

$$\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(V_1, V_2)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} \cdot F_{\frac{\alpha}{2}}(V_2, V_1)$$

$$F_{1-\alpha}(v_1, v_2) = \frac{1}{F_\alpha(v_2, v_1)}$$

$X_1, X_2 \sim$ normal. & independent.

$V_1 = n_1 - 1, \quad V_2 = n_2 - 1$

$n_A = 10, \quad n_B = 10, \quad S_A^2 = 5000^2, \quad S_B^2 = 6100^2$

$F_{\frac{10\%}{2}}(9, 9) = 3.18 \qquad \frac{1}{F_{5\%}(9,9)} = \frac{1}{3.18} = 0.314$

90% C.I. for $\frac{\sigma_A^2}{\sigma_B^2}$: $\left[\frac{S_A^2}{S_B^2} \cdot \frac{1}{F_{5\%}(9,9)},\ \frac{S_A^2}{S_B^2} \cdot F_{5\%}(9,9)\right] \rightarrow [0.211, 2.137]$

Thus, 90% C.I. for $\frac{\sigma_A}{\sigma_B}$: $[0.459, 1.462]$

8. A firm has been experimenting with two different physical arrangements of its assembly line. It has been determined that both arrangements yield approximately the same average number of finished units per day, so we are interested in the arrangement that has the smaller variance. Two independent random samples have been drawn, one for each arrangement. The data are:

Arr. 1	Arr. 2
$n_1 = 21$	$n_2 = 25$
$s_1^2 = 1408$	$s_2^2 = 3729$

Find a 90% and a 98% confidence interval for the ratio $\frac{\sigma_1^2}{\sigma_2^2}$.

90% C.I.

Recall: $\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(V_1, V_2)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} \cdot F_{\frac{\alpha}{2}}(V_2, V_1)$

$\alpha = 10\%$ $F_{5\%}(24, 20) = 2.08$ $\frac{1}{F_{5\%}(20, 24)} = \frac{1}{2.03}$

So. $\frac{1408}{3729} \cdot \frac{1}{2.03} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1408}{3729} \cdot 2.08$ $\frac{\sigma_1^2}{\sigma_2^2} \in [0.186, 0.785]$

98% C.I. $F_{0.01}(24, 20) = 2.86$ $\frac{1}{F_{0.01}(20, 24)} = \frac{1}{2.74}$

98% C.I. for $\frac{\sigma_1^2}{\sigma_2^2}$: $\left[\frac{1408}{3729} \cdot \frac{1}{2.74}, \frac{1408}{3729} \times 2.86\right] = [0.138, 1.080]$

9. Two different brands of latex paint are being considered for use. Drying time in hours is being measured on specimen samples of the use of the two paints. Sixteen specimens for each were selected and the drying times are as follows:

Paint A	3.5	2.7	3.9	4.2	3.6	2.7	3.3	5.2	4.2	2.9	4.4	5.2	4.0	4.1	3.4	3.7
Paint B	4.7	3.9	4.5	5.5	4.0	5.3	4.3	6.0	5.2	3.7	5.5	6.2	5.1	5.4	4.8	4.9

(a) Assume that the drying times are normally distributed with $\sigma_A = \sigma_B$. Find a 95% confidence interval for $\mu_B - \mu_A$ where μ_A and μ_B are the drying times

(b) construct a 90% confidence interval for σ_A^2/σ_B^2. Is the equal variance assumption valid?

(a) $\bar{X}_A = 3.8125$, $\bar{X}_B = 4.9375$. $n_A = n_B = 16$. $S_p = 0.7411$

$\bar{X}_B - \bar{X}_A \sim t\left(\mu_B - \mu_A, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\right)$

$\searrow$ σ is unknown.

$t_{2.5\%} = 2.042$ (degree = 30)

assume $\sigma_A = \sigma_B$.

$\downarrow$

95% C.I. for $\mu_B - \mu_A$:

$$\left[(\mu_B - \mu_A) - t_{2.5\%} \times S_p \times \sqrt{\frac{\sigma_P^2}{n_A} + \frac{\sigma_P^2}{n_B}}, (\mu_B - \mu_A) + t_{2.5\%} \times S_p \times \sqrt{\frac{\sigma_P^2}{n_A} + \frac{\sigma_P^2}{n_B}}\right]$$

$= [0.572, 1.678]$

(b) Recall: $\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(V_1, V_2)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} \cdot F_{\frac{\alpha}{2}}(V_2, V_1)$

$F_{0.05}(15,15) = 32.801$ $\qquad \dfrac{1}{F_{0.05}(15,15)} = \dfrac{1}{2.40}$

$S_A = 0.7535$

$S_B = 0.7284 = \dfrac{1}{n_B - 1} \cdot \sum_{i=1}^{n_B} (X_B^i - \bar{X}_B)^2$

$$90\%\ \text{C.I. for } \frac{\sigma_A^2}{\sigma_B^2} : \left[\frac{0.7535^2}{0.7284^2} \cdot \frac{1}{2.40} , \frac{0.7535^2}{0.7284^2} \times 2.40 \right]$$

$$= [0.4459, 2.5683]$$

The interval includes 1, so the assumption $\sigma_A = \sigma_B$ seems reasonable.

10. Two random samples are drawn from approximately normal distributions:

$$X_1, \ldots, X_7 = 10, 11, 6, 17, 10, 4, 12$$
$$Y_1, \ldots, Y_5 = 11, 3, 18, 8, 5$$

Calculate a 90% confidence interval for $\frac{\sigma_x^2}{\sigma_y^2}$.

$\bar{x} = 10$, $\bar{y} = 9$. $\quad n_X = 7$, $n_Y = 5$ $\langle n_X - 1 = 6, n_Y - 1 = 4 \rangle$

$$S_X = \frac{1}{7-1} \cdot \left[(10-10)^2 + (11-10)^2 + (6-10)^2 + (17-10)^2 + (4-10)^2 + (12-10)^2 \right]$$

$$= 17\tfrac{2}{3}$$

$$S_Y = \frac{1}{5-1} \left[(11-9)^2 + (3-9)^2 + (18-9)^2 + (8-9)^2 + (5-9)^2 \right] = 34.5$$

Recall: $\dfrac{S_1^2}{S_2^2} \cdot \dfrac{1}{F_{\frac{\alpha}{2}}(V_1, V_2)} \leq \dfrac{\sigma_1^2}{\sigma_2^2} \leq \dfrac{S_1^2}{S_2^2} \cdot F_{\frac{\alpha}{2}}(V_2, V_1)$

$\dfrac{1}{F_{5\%}(6,4)} = \dfrac{1}{6.16} = 0.1623$, $\quad F_{5\%}(4,6) = 4.53$.

$$90\%\ \text{C.I. for } \frac{\sigma_X^2}{\sigma_Y^2} : \left[\frac{S_X^2}{S_Y^2} \cdot \frac{1}{F_{5\%}(6,4)} , \frac{S_X^2}{S_Y^2} \cdot F_{5\%}(4,6) \right]$$

$$= [0.08313, 2.3197]$$