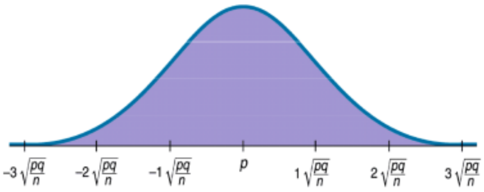
Week 6

Monday, 27 November 2023 9:11 AM

Sampling distribution Model :



$$N\left(p, \sqrt{\frac{pq}{n}}\right)$$

$p \rightarrow$ proportion of success
$n \rightarrow$ Sample sise

Normal model

$N(\mu, \sigma)$

$\mu \rightarrow$ Mean
$\sigma \rightarrow$ S.D.

$\mu = np$
$\sigma = \sqrt{npq}$

$\frac{np}{n} \rightarrow p$

$\frac{\sqrt{npq}}{n} \rightarrow \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$

Qualitative Data OR CATEGORICAL DATA $\} \rightarrow$ proportions

Example: 36% of the people are left handed

Quantitative Data OR Numerical Data $\} \rightarrow$ Mean and S.D.

Example: Heights of student in a class

171 cm, 162 cm, 180 cm, 168 cm

$\text{Mean} = \frac{\sum y}{n}$

$\text{S.D.} = \sqrt{\frac{(y-\bar{y})^2}{n-1}}$

1. Randomisation Condition: The sample should be a simple random sample of the population.
2. 10% Condition: If sampling has not been made with replacement, then the sample size, *n*, must be no larger than 10% of the population.
3. Success/Failure Condition: The sample size has to be big enough so that both *np* and *nq* are greater than 10.

The Central Limit Theorem (CLT)

The mean of a random sample has a sampling distribution whose shape can be approximated by a Normal model. The larger the sample, the better the approximation will be.

- For the sampling distributions for proportions and means, the standard deviations are based on population parameters
  - For proportions $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$
  - For means $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$

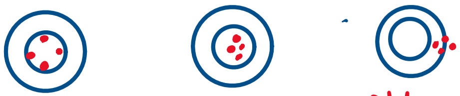
Example : Measuring the length of a calculator

12.4cm

12.35 , 12.45 , 12.39 , 12.42

Estimate ± Margin of Error

12.4 ± 0.05



Accuracy and Not precise

Accuracy and Precision

Not accurate and Precise

CONFIDENCE INTERVALS

Today's WEATHER

MAX - 20°c

99.7% → -5°c and 40°c

68% → 19°c and 22°c

95% → ✓ 15°c and 25°c

- The most commonly chosen confidence levels are 90%, 95%, and 98% (but any percentage can be used).

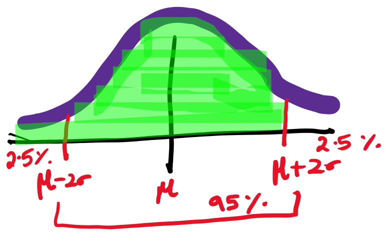
Confidence intervals for ONE PROPORTION :

Estimate ± Margin of Error

One proportion $\hat{p} \pm z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Critical value 90%, 95%, 98%.

	z^*
90%	1.645
95%	1.96
98%	2.33



95% + 2.5% = 97.5% → $\boxed{z = 1.96}$

Question 3 : Confidence interval example

If 11% of a random sample of 100 people have a particular gene, what can we say about the percentage of the population who have the gene?
Construct a 95% confidence interval for the actual percentage of the population who have the gene.

One proportion $\hat{p} \pm \overset{1.96}{z^*} \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$\hat{p} = 11\% = \frac{11}{100} = 0.11$

$\hat{q} = 89\% = 0.89$

$n = 100$

$= 0.11 \pm 1.96 \times \sqrt{\frac{0.11 \times 0.89}{100}}$

$= 0.11 \pm 1.96 \times 0.0312889$

$= 0.11 \pm 0.061$

(0.11 − 0.061) and (0.11 + 0.061)
Lower limit Upper limit

0.049 and 0.171

We are 95% Confident that between 4.9% and 17.1% of the people have this particular gene.

We are 95% Confident that between 4.9% and 17%
of the people have this particular gene.

Confidence Level:
0.95

Sample size (n):
100

Sample proportion ($\hat{p}$) or #successes:
0.11

Rounding:
4

Confidence interval: [0.04867 , 0.1713].

Success/Failure Condition:

$n\hat{p} > 10$ $\qquad$ $n\hat{q} > 10$

$100 \times 0.11 = 11 > 10$ $\qquad$ $100 \times 0.89 = 89 > 10$

2. From a survey of 283 randomly selected adults, 22% were smokers.
Calculate a 90% confidence interval for the true proportion of adults who smoke. (hint: remember to change z* from the value used in part 1)

$n = 283$

$z^*_{90\%} = 1.645$

$\hat{p} = 22\% = 0.22$

$\hat{q} = 78\% = 0.78$

One proportion $\quad \hat{p} \pm z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$= 0.22 \pm 1.645 \times \sqrt{\frac{0.22 \times 0.78}{283}}$

$= 0.22 \pm 0.04$

$= 0.18$ and 0.26

We are 90% Confident that the true proportion of adults who smoke is between 18% and 26%.

We can be 90% sure that the true proportion of adults who smoke is between 17.9% and 26.1%.

HYPOTHESES TESTING

→ Writing NULL and ALTERNATIVE HYPOTHESES

NULL → What we know

ALTERNATIVE → checking what we know

Example:

NULL $\longrightarrow$ H_0: There are 36% people who are left handed

ALTERNATIVE $\rightarrow$ H_A: 36% people are not left handed

OR

There are more than 36% people who are left handed

OR

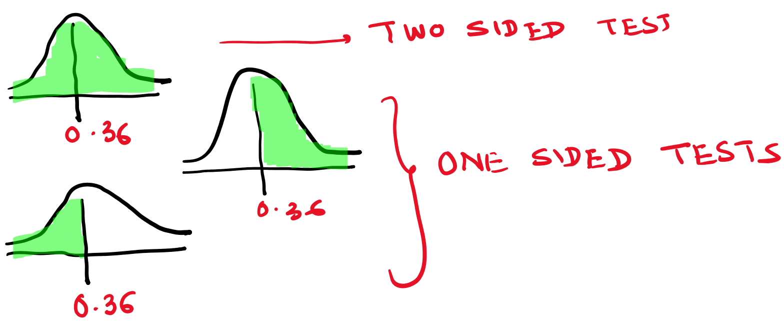
There are less than 36% people who are left handed

$H_0: p = 0.36$

$H_A: p \neq 0.36$

$p > 0.36$

$p < 0.36$



HYPOTHESES TESTING

Step 1: Write Null and Alternative Hypotheses

Step 2: Find the Test statistic.

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$ $\longrightarrow$ one proportion test statistic

Step 3: Z-score $\rightarrow$ Look up Normal table to find PROBABILITY VALUE [P-value]

Step 4: Compare P-value to ALPHA LEVELS (α)

10%, 5%, 1%

0.10 0.05 0.01

Step 5: CONCLUSION

P-value < Alpha level

P-value is LOW, Reject Null Hypotheses

P-value > Alpha level

P-value is HIGH, FAIL to Reject Null Hypothesis

Question 5 : Formulating hypotheses

Determine H_0 and H_A for the following:

1. Suppose that a pharmaceutical company is testing a new medication to see if side effects will be experienced by fewer than 20% of the patients who take it.

$H_0: p = 0.20$
$H_A: p < 0.20$

2. National data from the 1960s showed that 44% (p_0) of the adult population had never smoked. A more recent survey interviewed a random sample of 881 adults and found that 52% had never been smokers. Has the proportion of adults who have never smoked increased since the 1960s?

$H_0: p = 0.44$
$H_A: p > 0.44$

3. A vaccine states that it is 90% effective. A new vaccine is being trialed to see if it has a better or worse success rate. $H_0: p = 0.90$
$H_A: p \neq 0.90$

$<$: less than, fewer, Lower, decreased

$>$: more than, Higher, Increased

$\neq$: Changed, different, not similar

Question 6 : Hypothesis test example

National data from the 1960s showed that 44% of the adult population had never smoked. A more recent survey interviewed a random sample of 881 adults and found that 52% had never been smokers. Has the proportion of adults who have never smoked changed since the 1960s?

a) Create a 95% **confidence interval** for the current proportion of adults who have never smoked.
b) Carry out a **hypothesis test** to see if the survey data provide evidence that the proportion of adults who have never smoked has changed since the 1960's. Use α=0.05.

Orginal information

$p_0 = 0.44$
$q_0 = 0.56$
$n = 881$

Current study

$\hat{p} = 0.52$
$\hat{q} = 0.48$

(a)

95% → 1.96

$$\hat{p} \pm z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$= 0.52 \pm 1.96 \times \boxed{\sqrt{\frac{0.52 \times 0.48}{881}}}$$ → Standard Error

$$= 0.52 \pm 0.03299$$

$$= 0.487 \text{ and } 0.553$$

We are 95% confident that the current proportion of adults

$= 0.487$ and 0.553

We are 95% Confident that the current proportion of adults who have never smoked is between 48.7% and 55.3%.

Proportion Confidence Interval Calculator

Proportion confidence interval calculator with calculation steps, using the *normal distribution approximation (Wald interval)*, *binomial distribution*, and the *Wilson score interval*.

Confidence Level:	Sample size (n):
0.95	881
Sample proportion (p̂) or #successes:	Rounding:
0.52	3

Confidence interval: [0.487 , 0.553].

$$\hat{p} \pm z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.52 \pm 1.96 * \boxed{0.0168}$$ Standard error $\rightarrow \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$\underbrace{0.52}_{\text{Estimate}} \pm \underbrace{0.033}_{\text{Margin of Error}}$$

(b) $H_0 : p = 0.44$

$H_A : p \neq 0.44$

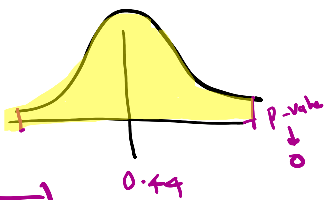
Changed $\longrightarrow \neq$

Test-statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.52 - 0.44}{\sqrt{\frac{0.44 \times 0.56}{881}}} = \frac{0.08}{0.0167}$$

$$Z = 4.79$$

P-value $\longrightarrow$ Look up Normal table



P-value = 0.000

$\boxed{\alpha = 0.05}$ — Given

P-value $<$ 0.05

P-value is LOW, REJECT NULL HYPOTHESES

Reject H_0 - the proportion of adults who have never smoked has changed significantly.

OK

The proportion of adults who have never smoked are not 44%.

One Sample Proportion Test

Proportion Z-test and Binomial test

Video Two sample proportion calculator

Tails: Two ($H_1: p \neq p_0$)	Digits: 6
Significance level (α): 0.05	Continuity: True
h effect size:	Calculate the expected h effect size: Calculate h
Name: people who never smoked	Expected proportion (P_0): 0.44
Proportion ($\hat{p}$) or total number (x): 0.52	Sample size (n): 881

The test statistic Z equals **4.749694**,

The p-value equals **0.00000203725**,

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Since p-value < α, H_0 is rejected.

The proportion ($\hat{p}$) of **people who never smoked's** population is considered to be **not equal to** the expected proportion (P_0).

In other words, the difference between the sample proportion ($\hat{p}$) of **people who never smoked** and expected proportion (P_0) is big enough to be statistically significant.

→ You have found new information OR Reject Null Hypothesis

Question 7 : Hypothesis test activity

Suppose that a pharmaceutical company claims that side effects will be experienced by fewer than 20% of the patients who use a particular medication.

In a clinical trial with 400 patients, 68 patients experienced side effects.

1. Is there evidence that the company's claim is true? Perform a hypothesis test using α = 0.05.

$p_0 = 0.20$

$q_0 = 0.80$

Current study

$n = 400$

$\hat{p} = \frac{68}{400} = 0.17$

$\hat{q} = 0.83$

$H_0: p = 0.20$

$H_A: p < 0.20$

HYPOTHESES - H_0, H_A ✓

TEST STATISTIC ✓

P-VALUE ✓

$$Z = \frac{\hat{p} - p_0}{\quad} = \frac{0.17 - 0.20}{\quad} = \frac{-0.03}{\quad}$$

TEST STATISTIC ✓
P-VALUE ✓
Comparison ✓
Conclusion ✓

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.17 - 0.20}{\sqrt{\frac{0.20 \times 0.80}{400}}} = \frac{-0.03}{0.02}$$

$\boxed{Z = -1.5}$

9	8	7	6	5	4	3	2	1	0	
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	−1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	−1.5

P-value = 0.0668 $\alpha = 0.05$

P-value > 0.05

P-value is HIGH, FAIL to Reject Null Hypotheses

∴ 20% of the patients will experience side effects
OR
COMPANY's claim is not true, since we fail to Reject H_0
OR
Results are not statistically significant as we fail to Reject H_0.

Since the P-value is >0.05, fail to reject H_0.
At a 5% significance level, there is insufficient evidence that fewer than 20% of patients using the medication experience side effects.

One sided test, P-value = 0.0668
Two sided test, P-value = 0.0668 × 2 = 0.1336

JUST A NOTE

2.7 A random sample of 262 people tried a special exercise routine that lasted 2 months to lose weight. The weights of these people were measured both at the beginning and the end of the routine. Out of them, 226 individuals experienced a weight loss of at least 5%. Is there evidence that this exercise results in an average weight loss of at least 5% in more than 81% of individuals?

Current study
$n = 262$
$\hat{p} = \frac{226}{262} = 0.86$
$\hat{q} = 0.14$

$p_0 = 0.81$
$q_0 = 0.19$

$H_0: p = 0.81$
$H_A: p > 0.81$

a) Write appropriate hypotheses.
b) Check the assumptions and conditions. Randomisation, 10% Condition OR Success/Failure Condition
c) Perform the hypothesis test (state $\hat{p}$, $\hat{q}$, z-score and P-value).
d) State your conclusion. Use a significance level of 5%. Comparison and Conclusion $\alpha = 0.05$
e) Give a 90% confidence interval for the true proportion of people who may have an average weight loss of at least 5% by this exercise and interpret your interval (state also Critical value z*, formula and value for Standard Error, and formula for the confidence interval).

Z-Score

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0}}$$

Look up Normal table for P-value

Z-score

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Look up Normal table for P-value

Confidence interval:

$$\hat{p} \pm z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

standard error

$z^*(90\%) = 1.645$