Sampling Distributions and Standard Errors

Lectures 15–17

QBA201 – Quantitative Business Analysis

Shi Qiu – Fall 2026

Recall: daily demand and average demand differ

Lecture 14: petrol demand on one day is normal, with mean 1,000 litres and SD 100 litres.

Compare one day with the average of 25 independent days.

Discuss: Which is more likely to exceed 1,100 litres? Explain before calculating.

One day uses SD 100. The average uses standard error $100/\sqrt{25} = 20$.

Draw a random sample from a customer register

A constructed register contains 24 customer orders. Amounts are in AED.
Select 4 distinct customer IDs at random, without replacement.
Calculate one mean from those 4 orders.

20	40	60	80	100	120
20	40	60	80	100	120
20	40	60	80	100	120
20	40	60	80	100	120

Population mean: 70 AED.

Discuss: Must a correctly selected sample also average exactly 70 AED?

Open the simulation: redraw the sample and watch the selected IDs and mean.

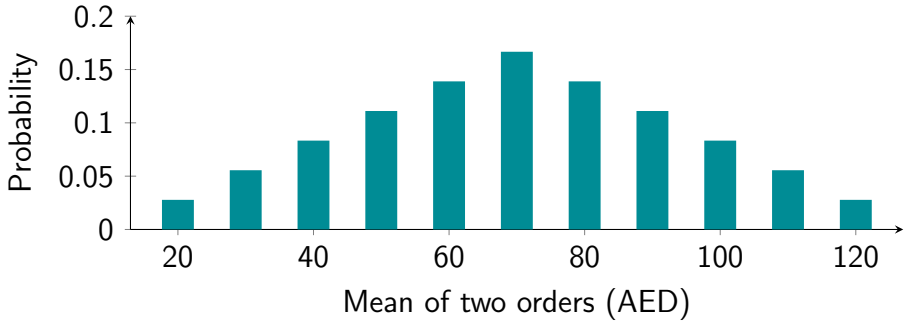
Two independent orders give 36 possible samples

For a separate model, each order is equally likely to be 20, 40, 60, 80, 100 or 120 AED. Draw independently, with replacement.

Ordered sample (AED)	Probability	Sample mean (AED)
$(20, 20)$	$1/36$	20
$(20, 40)$	$1/36$	30
$(40, 20)$	$1/36$	30
$(120, 120)$	$1/36$	120

Explore all 36 pairs. A point in the sampling distribution is a whole sample's mean.

More pairs produce a central average



Mean: 70 AED. Probability of a 70 AED average: 6/36.

Discuss: Why is a mean of 70 more likely than a mean of 20?

Averaging preserves the center and reduces spread

Quantity	One order	Mean of two orders
Mean (AED)	70	70
Variance (AED^2)	1,166.6667	583.3333
SD (AED)	34.1565	24.1523

The orders are independent and follow the same six-value model.

The variance halves. The SD is divided by $\sqrt{2}$, not by 2.

The mean and standard error of an average

Let $\bar{X}$ be the mean of n independent orders from the same population, with population mean μ and SD $\sigma > 0$.

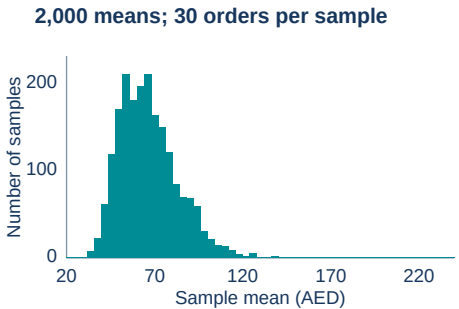
$$\mathrm{E}(\bar{X}) = \mu, \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

$$\mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

The population SD describes individuals. The standard error describes sample means. Both are measured in AED here.

Build the sampling distribution: then change n

Most orders are small; a few are large.
Draw n independent orders. Record their mean. Repeat.
Compare $n = 1, 5, 30, 100$ in the simulation.



For independent draws from the same population with finite, positive variance:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$ approaches a standard normal distribution as n grows.

Exact normality and approximate normality

Normal population

An average of independent normal observations is exactly normal for any n.

Nonnormal population

The normal approximation can improve with larger n under the CLT conditions.

Strong skewness or extreme tails can require larger samples.

Standardize using the statistic's spread

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

1. Identify whether the event concerns one observation or an average.
2. Find the relevant SD or standard error.
3. Standardize the stated cutoff.

A probability about average warehouse task time

Task time is normal: mean $\mu = 80$ minutes, SD $\sigma = 12$ minutes. Observe $n = 36$ independent tasks.

$$\mathrm{SE}(\bar{X}) = 12/\sqrt{36} = 2 \text{ minutes.}$$

$$\mathrm{P}(\bar{X} > 84) = \mathrm{P}\left(Z > \frac{84 - 80}{2}\right) = 1 - \Phi(2) = 0.0228.$$

The population is normal, so the mean is exactly normal. A CLT approximation is unnecessary here.

A finite population can reduce the standard error

For simple random sampling without replacement from N units:

$$\mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

Here σ uses the population-variance divisor N.

Observing a substantial share leaves less sampling uncertainty. The course usually omits this correction when $n/N \leq 0.05$.

The sample proportion is a sample average

Let X count successes among n binary observations.

$$\hat{p} = X/n.$$

Equivalently, code each success as 1 and failure as 0. Then $\hat{p}$ is the average of those indicators.

Discuss: If 72 of 120 sampled customers buy, what is $\hat{p}$?

Keep a purchase count separate from a share

Twenty independent customer opportunities each have purchase probability 0.5. Let X be the number of purchases.

$$X \sim \text{Bin}(20, 0.5), \qquad \text{E}(X) = 10.$$

$$\text{Var}(X) = 5, \qquad SD(X) = \sqrt{5} = 2.2361.$$

Discuss: For the sample share $\hat{p} = X/20$, what happens to the mean and SD?

$\text{E}(\hat{p}) = 0.5$ and $SD(\hat{p}) = 2.2361/20 = 0.1118$.

Check the success and failure counts

The course uses:

$$np \geq 5, \qquad n(1-p) \geq 5.$$

For $n = 20, p = 0.5$, both expected counts equal 10.

For $n = 200, p = 0.01$, $np = 2$, so this rule fails.

A large n alone is not enough when one outcome is rare.

Continuity correction matches one integer bar

For a normal approximation Y to the binomial count X:

$$\mathrm{P}(X = 10) \approx \mathrm{P}(9.5 < Y < 10.5).$$

The interval covers the width of the discrete count's bar.

Use continuity correction when the question calls for it. Follow explicit instructions for sample-proportion approximations.

Check approximation against the exact probability

$$\mathrm{P}(X = 10) = \binom{20}{10}(0.5)^{20} = 0.1762.$$

For Y normal with mean 10 and SD $\sqrt{5}$:

$$\mathrm{P}(9.5 < Y < 10.5) = 0.1769.$$

Mean and standard error of a sample proportion

Under independent Bernoulli sampling:

$$\mathrm{E}(\hat{p}) = p, \qquad \mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}.$$

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}.$$

The probability calculation uses the given population proportion p, not an estimated substitute unless the question asks for one.

Constructed practice: a proportion above a cutoff

Assume $p = 0.60$, $n = 400$, and independent binary observations.

$$np = 240, \quad n(1-p) = 160, \quad \text{SE} = \sqrt{.6(.4)/400} = 0.024495.$$

Without continuity correction:

$$\text{P}(\hat{p} > .65) \approx 1 - \Phi(2.0412) = 0.0206.$$

Two sample means need two sampling models

Take independent random samples from two populations.

If both populations are normal, $\bar{X}_1 - \bar{X}_2$ is normal.

Otherwise a normal approximation can be justified when both sample means are well approximated by normal distributions.

Independent groups and paired observations need different calculations.

Center and spread of the difference

$$\mathrm{E}(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2.$$

For independent samples:

$$\mathrm{SE}(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

The variances add even though the statistic is a difference.

A training comparison is not automatically an effect

A company compares mean assessment scores from two employee training groups.

Differences can reflect sampling variation, prior preparation, selection, measurement or instruction.

Discuss: What information would make the comparison more credible?

A sampling model describes variation. It does not by itself establish a causal training effect.

Two employers: specify the salary models

	Employer 1	Employer 2
Mean annual salary (euros)	62,000	60,000
Population SD (euros)	14,500	18,300
Independent sample size	50	60

Assume normal salary models and independent random samples.

Discuss: How likely is the first sample's mean salary to exceed the second's?

Calculate the probability of a positive observed difference

$$D = \bar{X}_1 - \bar{X}_2, \quad \mathrm{E}(D) = 2,000, \quad \mathrm{SE}(D) = 3,128.3382.$$

$$\mathrm{P}(D > 0) = \mathrm{P}\left(Z > \frac{0 - 2,000}{3,128.3382}\right) = 0.7387.$$

Inference starts with an unknown population quantity

In practice, μ or p is usually unknown.

We observe a sample statistic, then use its sampling behavior to assess plausible population values.

A sampling distribution connects the observed statistic to the unknown parameter.