POLS0008 (Week 5): Normal Distribution, CLT, Standard Error

Breakout Seminar Review + Problems

James Rice

Thursday, February 12, 2026

Today: what we're doing and why

- **Big picture:** we want to make claims about a population using a sample.
- **Problem:** different random samples give different answers (sampling variation).
- **Week 5 recap:** normal distribution, z-scores, CLT, and SE as a measure of precision.
- **Seminar activity:** work through the seminar questions (first answer *questions* in groups, then review *answers* as a class).

Key takeaway to keep in mind

SD describes spread of *data*; SE describes spread of an *estimate across repeated samples*.

Why inference needs probability

- We usually want to learn about a **population** (all voters, all students, all households).
- But we only observe a **sample**: a limited, noisy snapshot.
- So: two samples from the same population can give different answers.

Goal of Week 5

Use probability to quantify uncertainty in sample-based estimates (means, proportions).

Core idea

Inference is about **how estimates vary under repeated random sampling**.

Point estimates and uncertainty

- A **point estimate** is a single-number guess of a parameter:

$$\bar{x} \text{ estimates } \mu, \qquad \hat{p} \text{ estimates } p. \tag{1}$$

- But a point estimate alone does not tell us *how precise* it is.
- Precision depends on:
  - underlying variability in the population (SD), and
  - the sample size n (more data $\Rightarrow$ more precision).

The quantity that captures the precision of an estimate (how much it would vary across repeated samples) is the **standard error (SE)**.

Standard error of the mean

$$\mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} \quad \text{(often estimated by } s/\sqrt{n}\text{).}$$

Variance and standard deviation (spread)

Definitions

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2], \qquad \mathrm{SD}(X) = \sigma = \sqrt{\mathrm{Var}(X)}.$$

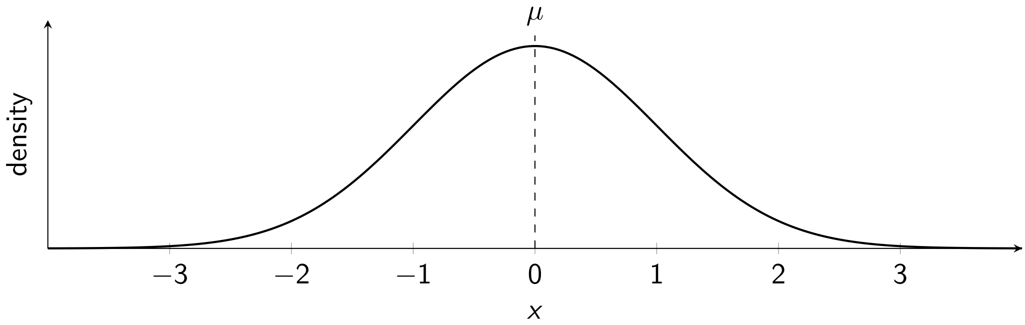
- Variance: squared spread (useful mathematically, less interpretable).
- Standard deviation: typical distance from the mean (in original units).
- If SD is large, values are dispersed; if SD is small, values cluster tightly.

Why it matters

SD is the scale we use to compare deviations from the mean (e.g., via z-scores).

Normal distribution: shape + parameters

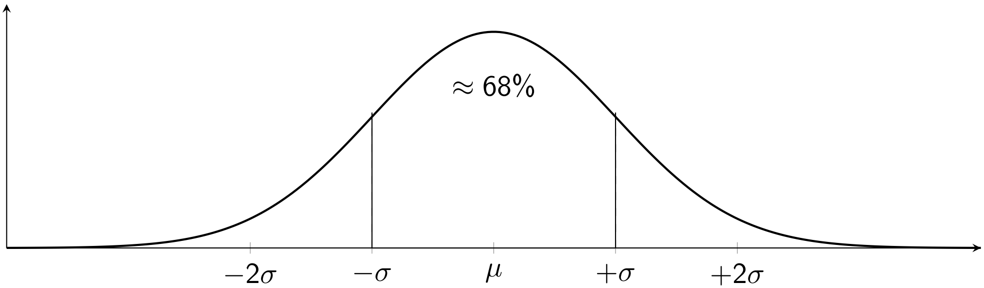
$$X \sim \mathcal{N}(\mu, \sigma^2), \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$



- μ shifts the curve left/right; σ stretches/compresses it.
- Areas under the curve correspond to probabilities (total area $= 1$).

The 68–95 rule (quick calibration)

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68, \qquad P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95.$$



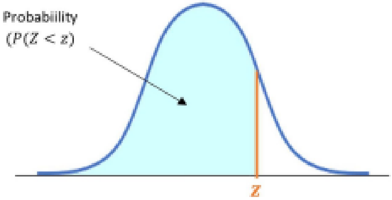
- Think of σ as a *typical step size* away from the average μ: most observations aren't far from μ.
- The area under the curve is probability, so "$\pm 1\sigma$" means: about 68% of values fall in the central part of the bell.
- Expanding to "$\pm 2\sigma$" captures about 95% of values.

Z-scores and normal probabilities

Standardise to the **standard normal**:

$$Z \sim \mathcal{N}(0,1), \qquad z = \frac{x-\mu}{\sigma}.$$

- A z-score turns a raw value (e.g., $\bar{x}$) into a comparable metric.
- Tables usually report $\Phi(z) = P(Z \leq z)$ (area to the left).
- Upper tail: $P(Z \geq z) = 1 - \Phi(z)$.



CLT and standard error (mean)

Sampling distribution of the mean:

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

CLT (informal)

For large enough n,

$$\bar{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

even if the original data are not exactly normal.

Seminar questions (1/2)

Q1. In your own words, what is the *sampling distribution of means*?

Q2. In your own words, what is *standard error*?

Q3. How does the standard deviation help us arrive at a *z-score*? What is a z-score?

Seminar answers (1/2)

A1. Sampling distribution of means

Repeatedly sample size n, compute $\bar{x}$ each time; the distribution of those $\bar{x}$'s is the sampling distribution. It is centered on μ: $\mathbb{E}(\bar{X}) = \mu$.

A2. Standard error

Typical spread of a statistic across repeated samples. For the mean:

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

Bigger $n \Rightarrow$ smaller SE.

A3. Z-score

$$z = \frac{x - \mu}{\sigma}$$

$=$ number of SDs x is from the mean.

Seminar questions (2/2)

Consider a sample of $n = 205$ students with mean exam score $\bar{x} = 61$ and SD $s = 18$.

Q4. Compute the *standard error* of the mean.

Q5. Compute the z-score for an exam score $x = 72$. Then find the proportion scoring ≥ 72.

Q6. What is the relationship between sample size and standard error? How does this connect to the CLT?

Z-score lookup table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7996	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

- Find your z-score by matching the *row* (first digit + first decimal, e.g. 0.6) and the *column* (second decimal, e.g. 0.01); the entry is $\Phi(z) = P(Z \leq z)$ (area to the left).
- Convert as needed: $P(Z \geq z) = 1 - \Phi(z)$.

A4. Standard error (given $s = 18, n = 205$)

$$\mathrm{SE}(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{18}{\sqrt{205}} \approx \frac{18}{14.318} \approx 1.257.$$

A5. Z-score and proportion ≥ 72

$$z = \frac{72 - 61}{18} = \frac{11}{18} \approx 0.611.$$

$$P(X \geq 72) = P(Z \geq 0.611) = 1 - \Phi(0.611) \approx 1 - 0.7291 = 0.2709 \approx 27\%.$$

A6. n, SE, and CLT

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

Wrap-up

- The normal curve is our "probability map": μ sets the centre, σ sets how spread out the data are.
- A z-score just rescales a value into "how many SDs from the mean?", so we can read off probabilities from the Z-table.
- The CLT explains why means behave nicely: if you take enough observations, the *distribution of sample means* is close to normal.
- Standard error is the spread of those sample means: $\mathrm{SE}(\bar{X}) = \sigma/\sqrt{n}$. Bigger $n \Rightarrow$ smaller SE $\Rightarrow$ more precise estimates.