Today, we will:

- Do a review example problem – PDFs
- Review the pdf module: **The Gaussian or Normal Probability Density Function**
- Do some example problems – Gaussian PDFs and probability estimates

Example: PDFs

Given: A sample consists of 1000 length measurements. The sample mean is 53.65 cm and the sample standard deviation is 1.25 cm. 136 measurements lie between 55.0 < x ≤ 57.0 cm.

To do:

(***a***) Estimate the probability (%) that a length measurement lies between 55.0 and 57.0 cm.
(***b***) Calculate the transformed variables z_1 at x_1 = 55.0 cm and z_2 at x_2 = 57.0 cm.
(***c***) Discuss why it is useful to transform from $f(x)$ to $f(z)$.

Solution:

(a) $\text{Probability} = \dfrac{\text{Frequency}}{n} = \dfrac{\text{\# readings in this bin}}{\text{total \# readings}} = \dfrac{136}{1000} = 0.136 = \boxed{13.6\%}$

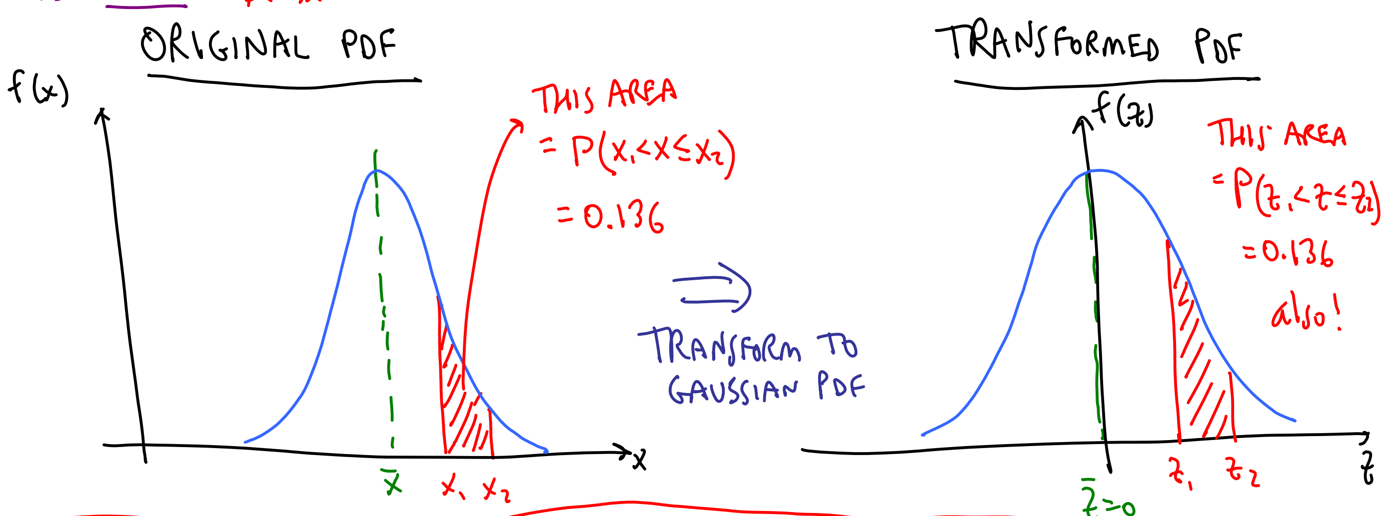
We write this as $\boxed{P(55.0 < X \le 57.0) = 0.136 \text{ or } 13.6\%}$

(b) $z_1 = \dfrac{x_1 - \mu}{\sigma} \approx \dfrac{x_1 - \bar{x}}{S}$ (Since we don't know μ or σ, we approximate) $= \dfrac{(55.0 - 53.65)\cancel{\text{cm}}}{1.25\ \cancel{\text{cm}}} = \boxed{1.08 = z_1}$

Notice → x & S have same units & z is dimensionless

$z_2 = \dfrac{x_2 - \mu}{\sigma} \approx \dfrac{x_2 - \bar{x}}{S} = \dfrac{57.0 - 53.65}{1.25} = 2.68 \qquad \boxed{z_2 = 2.68}$

(c) Discuss: ☆☆ THIS IS VERY IMPORTANT ! ☆☆



☆☆ $\boxed{\text{In general, } P(x_1 < x \le x_2) = \int_{x_1}^{x_2} f(x)\,dx = P(z_1 < z \le z_2) = \int_{z_1}^{z_2} f(z)\,dz}$ ☆☆

THE PROBABILITIES ARE THE SAME IN EITHER PDF (ORIGINAL or TRANSFORMED)

Example: Confidence level

Given: Many voltage readings are taken from a power supply.

- The sample mean voltage reading is $\overline{V} = 10.12$ V
- The sample standard deviation of all the readings is 0.022 V
- We assume that the errors in the readings are purely random

To do: Write the voltage to 95% confidence level.

Solution: In other words, we want to write $V = \overline{V} \pm$ ___ Volts

- For standard 95% confidence level, we want $\pm 2\sigma \approx \pm 2S$

[we don't know the population std. dev., so we approximate with the sample std. dev. $\rightarrow \sigma \simeq S$.]

- Here $\pm 2S = \pm 2(0.022 \text{ V}) = \pm 0.044$ V, $\therefore$ $\boxed{V = 10.12 \pm 0.044 \text{ V}}$

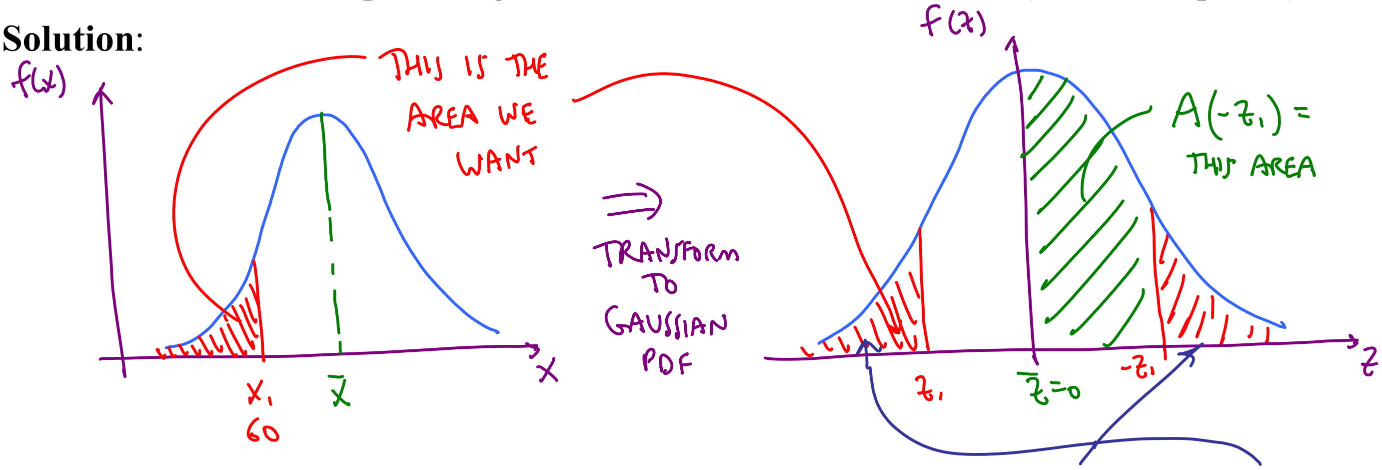
Note: It is commonly accepted to write the confidence interval

Example: Probability – exam scores

Given: In one of Professor Cimbala's midterm exams, the mean was 73.6 (out of 100 possible points) and the standard deviation was 9.2. We assume that the distribution of exam scores is Gaussian. The cutoff grade for a D is 60 points.

To do: Predict the percentage of students who failed the exam (score < 60 points).

Solution:



BUT, SINCE THE GAUSSIAN PDF IS SYMMETRIC, THIS AREA = THIS AREA

THE PROBABILITY IS $P(x < 60) = P(z > -z_1) = 0.5 - A(-z_1)$

$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{60 - 73.6}{9.2} = -1.478 \rightarrow A(-z) = A(z)$. Here, look up $A(1.478)$ in table $\rightarrow$ $A(-z_1) = 0.43029$

So, finally, $P(x < 60) = \frac{1}{2} - A(-z_1) = \frac{1}{2} - 0.43209 = 0.06971$ or $\boxed{\approx 7\%}$

Example: Probability

Given: 100 velocity measurements are taken in a wind tunnel.

- The sample mean velocity is $\bar{V} = 5.126$ m/s
- The sample standard deviation of all the readings is 0.0690 m/s
- We assume that the errors in the readings are purely random

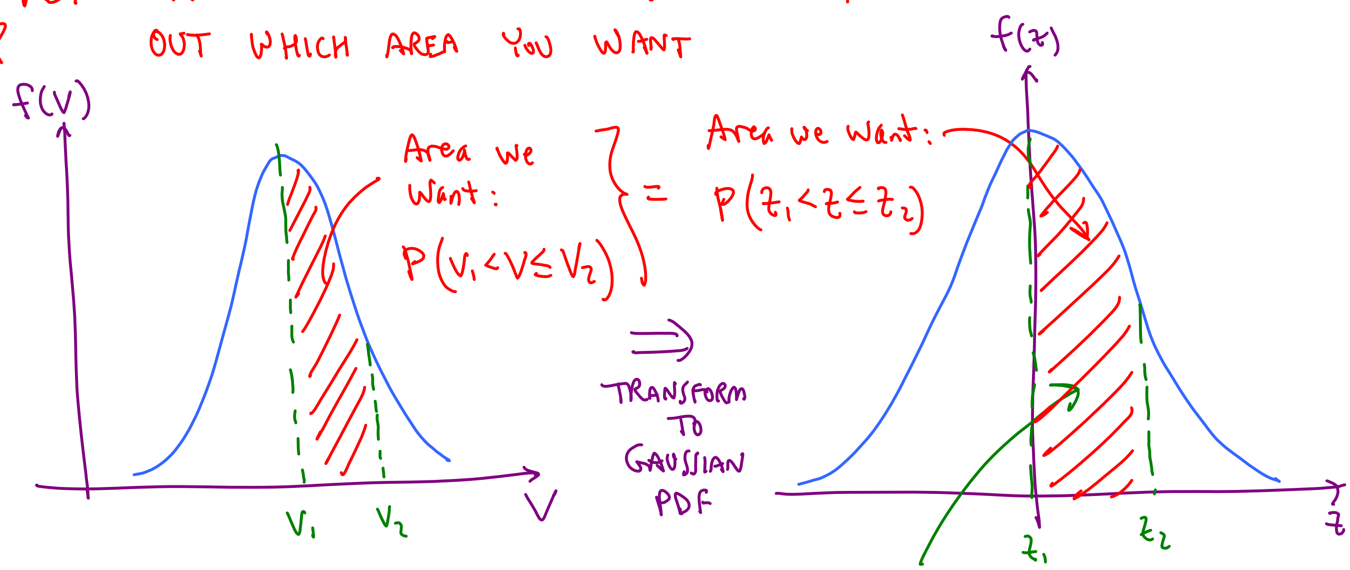
(*a*) To do: Calculate the probability that the velocity of a random measurement is in the range 5.126 m/s < *V* < 5.200 m/s. In other words, calculate *P*(5.126 < *V* < 5.200).

(*b*) To do: Calculate the probability that the velocity of a random measurement is in the range 5.000 m/s < *V* < 5.200 m/s. In other words, calculate *P*(5.000 < *V* < 5.200).

Solution: let $z_1 = \frac{V_1 - \mu}{\sigma} \approx \frac{V_1 - \bar{V}}{S} = \frac{5.126 - 5.126 \text{ m/s}}{0.0690 \text{ m/s}} = 0 = z_1$

Similarly, $z_2 = \frac{V_2 - \bar{V}}{S} = \frac{5.200 - 5.126}{0.0690} = 1.072 = z_2$

KEY: DRAW SKETCH OF BOTH PDFs (ORIGINAL & TRANSFORMED) & FIGURE OUT WHICH AREA YOU WANT



THIS AREA IS ALSO, BY DEFINITION = $A(z_2)$
(since $z_1 = 0$ here)

So, $P(V_1 < V \le V_2) = P(z_1 < z \le z_2) = A(z_2) = A(1.072) = 0.3581$

(interpolate from the $A(z)$ table)

$$\boxed{\text{Finally, } P(5.126 < z \le 5.200) = 0.3581 \text{ or } 35.8\%}$$

(b) Calculate $P(5.000 < V \leq 5.200)$

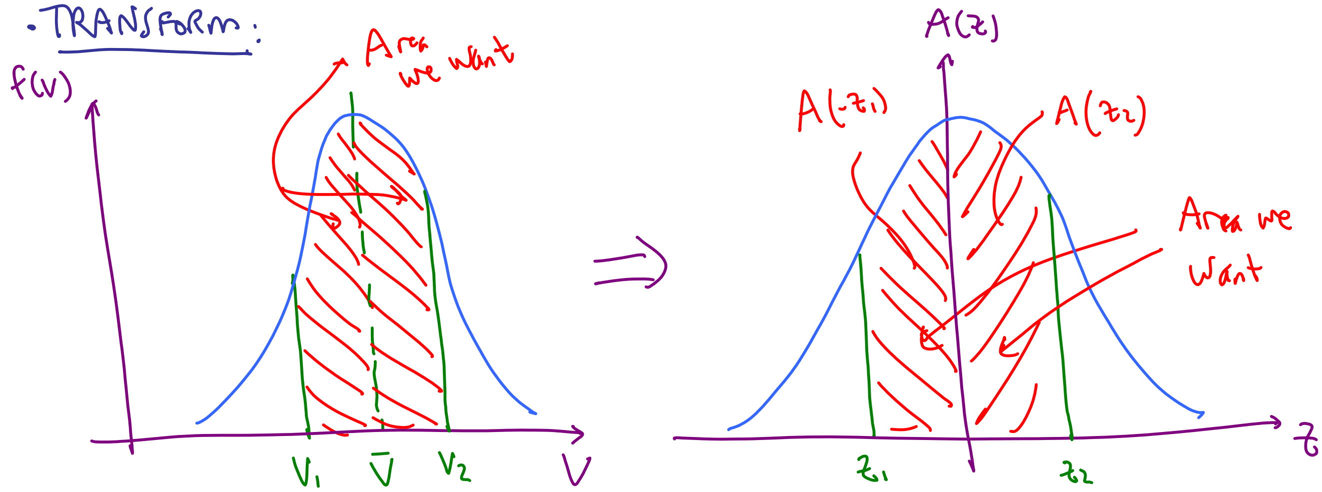
- Let $x_1 = 5.000 \rightarrow z_1 \approx \dfrac{V_1 - \bar{V}}{S} = \dfrac{5.000 - 5.126}{0.0690} = \underline{-1.826 = z_1}$

- Similarly, @ $X_2 = 5.200$, $\underline{z_2 = 1.072}$ (same as in Part (a))

- Table $\rightarrow$ $A(z_1) = A(-z_1) = A(1.826) = \underline{0.4661}$

(interpolate) $A(z_2) = A(1.072) = \underline{0.3581}$

- TRANSFORM:



- Here, the area we want is both red shaded areas, so...

$$P(V_1 < V \leq V_2) = P(z_1 < z \leq z_2) = A(-z_1) + A(z_2) = A(z_1) + A(z_2)$$

$$= 0.4661 + 0.3581 = 0.8242$$

$$\therefore \boxed{P(5.000 < V \leq 5.200) = 0.8242 \text{ or } 82.4\%}$$

Example: PDFs

Given: A sample consists of 400 voltage measurements. The sample mean is 35.52 V and the sample standard deviation is 1.84 V. There are 7 measurements in the bin $39.0 < x \le 39.5$ V.

To do:

(***a***) Estimate the probability (%) that a voltage measurement lies between 39.0 and 39.5 V.

(***b***) Calculate the transformed variable z at the midpoint of this bin, i.e., at $x = 39.25$ V.

(***c***) Calculate $f(z)$ at $x = 39.25$ V.

(***d***) Compare the answer to Part (c) with the analytical prediction of $f(z)$ at $x = 39.25$ V for a Gaussian (normal) pdf.

Solution:

(a) $\text{Prob.} = \dfrac{\text{Frequency}}{n} = \dfrac{7}{400} = 0.0175 \rightarrow \boxed{1.75\%}$

(b) $z = \dfrac{x-\mu}{\sigma} \approx \dfrac{x-\bar{x}}{S} = \dfrac{39.25 - 35.52\ \cancel{V}}{1.84\ \cancel{V}} = 2.027 \rightarrow \boxed{z = 2.027}$

(c) $f(z) = \sigma f(x)$, but $f(x) = \dfrac{\text{Probability}}{\Delta x}$, so ... $\underline{f(z) \approx S \dfrac{\text{Prob}}{\Delta x}}$

#'s $\rightarrow f(z) = (1.84\ V)\dfrac{0.0175}{0.5000\ V} = 0.0644$ $\boxed{f(z) = 0.0644}$

Note: Δx for this bin from 39 to 39.5 is 0.5, & we assume this is "exact" → lots of significant digits.

(d) Analytical $f(z) \rightarrow$ $f(z) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{z^2}{2}\right)$ for a Gaussian PDF.

Plug in $z = 2.027$, get $\boxed{f(z)_{\text{analytical}} = 0.0511}$

Compare: $\left.\begin{array}{l} f(z)_{\text{experimental}} = 0.0644 \\ f(z)_{\text{analytical}} = 0.0511 \end{array}\right\}$ Reasonably close but not exact

Why the disagreement?
- 400 pts is not ∞ → need more data
- errors may not be purely random, and so the PDF may not be perfectly Gaussian