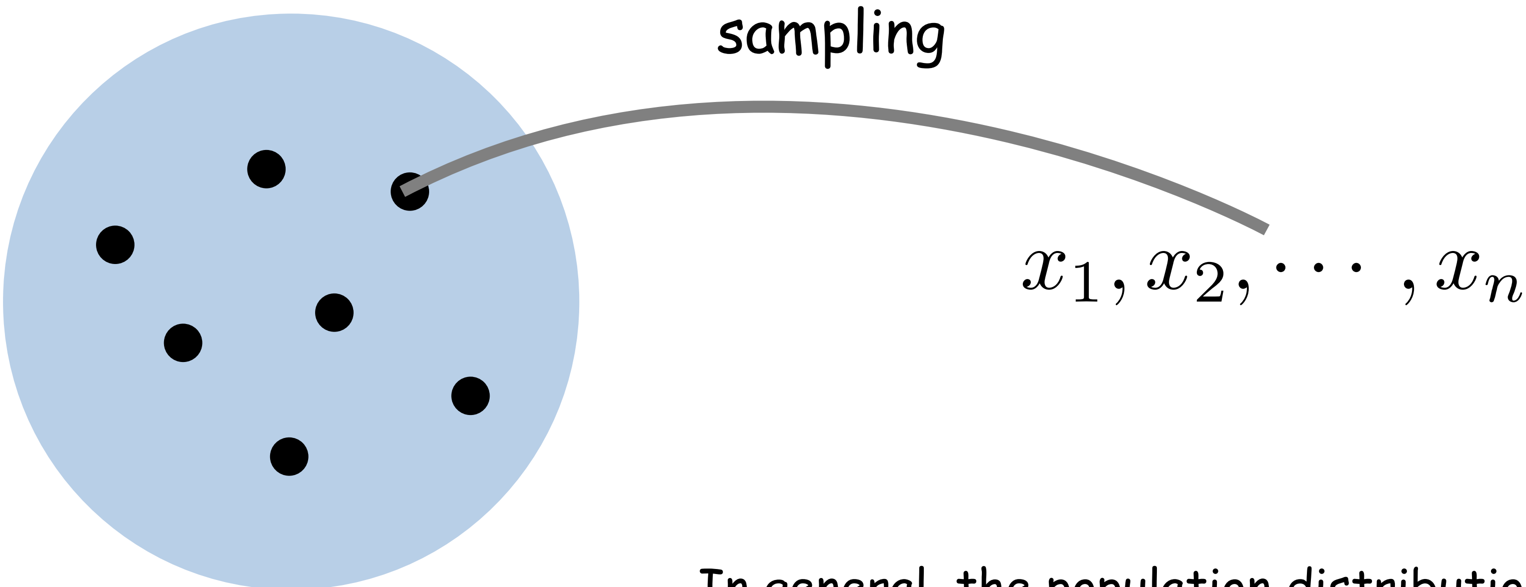
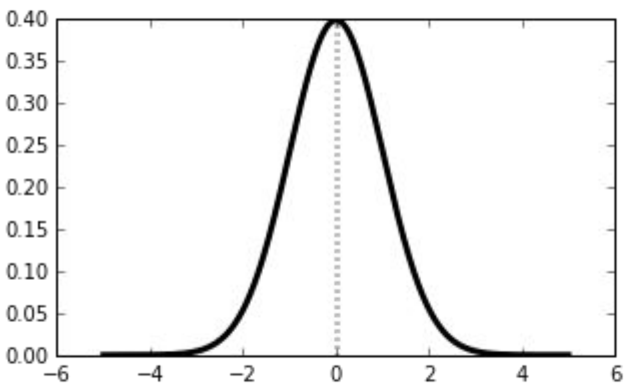
통계분석
Statistical Analysis

Probability Plot

Sampling from Population Distribution





Population distribution

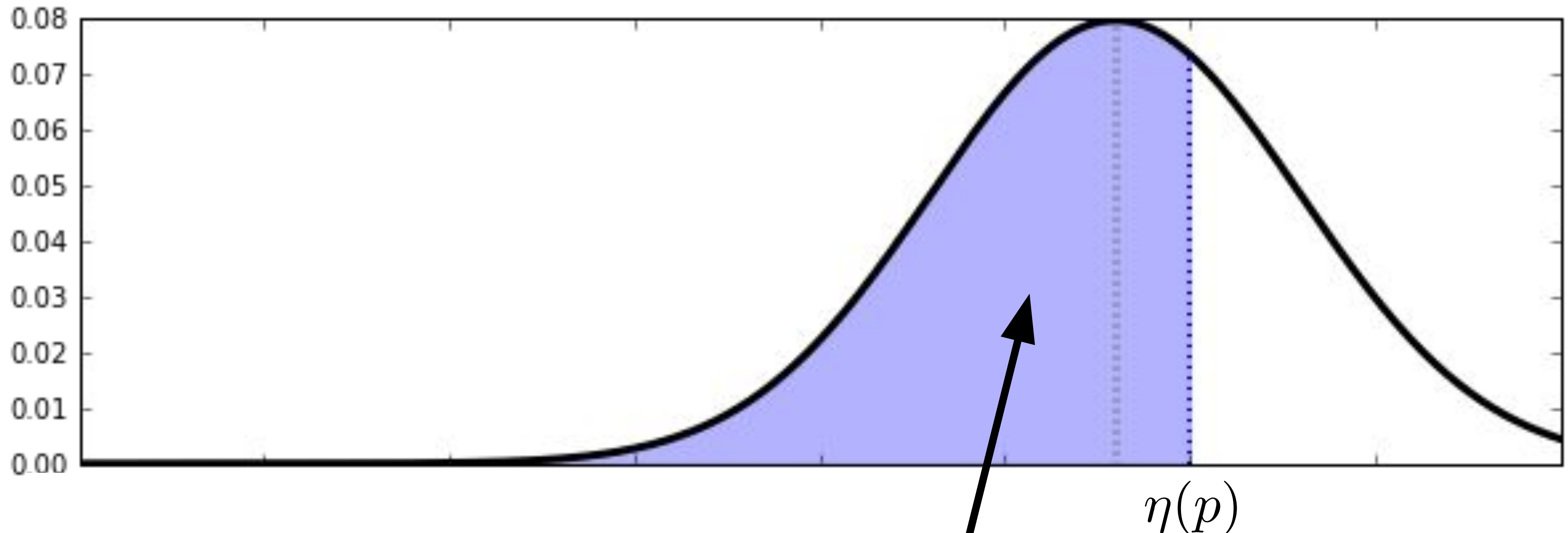
In general, the population distribution is unknown.

Statistical analysis (e.g. statistical inferences) requires to know the underlying distribution of population.

→ We can figure out the underlying distribution that the sample data is based on, by using the probability plot.

Review: Percentiles of Distributions

$(100p)$th percentile of some continuous distribution $= \eta(p)$



$$\int_{-\infty}^{\eta(p)} f(x)dx = p$$

Sample Percentiles

Can we define percentiles for sample data?

$$x_1 < x_2 < x_3 < x_4 < x_5 < x_6 < x_7 < x_8 < x_9 < x_{10}$$

Sample Percentiles

Can we define percentiles for sample data?

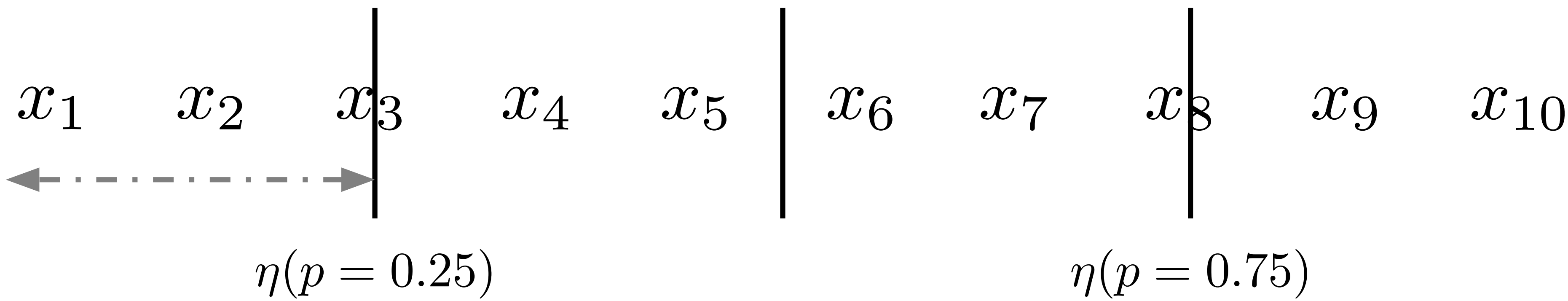
$$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \;\Big|\; x_6 \quad x_7 \quad x_8 \quad x_9 \quad x_{10}$$

$$\frac{5}{10} = 0.5$$

$$\eta(p = 0.5)$$

Sample Percentiles

Can we define percentiles for sample data?



Sample Percentiles

Order the *n* sample observations from smallest to largest. Then the ith smallest observation in the list is taken to be the ***[100(i-0.5)/n]***th sample percentile.

$$\begin{array}{cccccccccc}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\eta(p = .05) & \eta(.15) & \eta(.25) & \eta(.35) & \eta(.45) & \eta(.55) & \eta(.65) & \eta(.75) & \eta(.85) & \eta(.95)
\end{array}$$

Sample Percentiles

Order the *n* sample observations from smallest to largest. Then the ith smallest observation in the list is taken to be the ***[100(i-0.5)/n]***th sample percentile.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\eta(p=.05)$	$\eta(.15)$	$\eta(.25)$	$\eta(.35)$	$\eta(.45)$	$\eta(.55)$	$\eta(.65)$	$\eta(.75)$	$\eta(.85)$	$\eta(.95)$

If the sample data comes from some specifying distribution of population, It is plausible that sample percentiles are quite close to percentiles of the underlying distribution.

1. Assume one possible distribution of population.
2. Calculate percentiles of the distribution corresponding to sample percentiles
3. Compare sample data and percentiles calculated in the previous step.
4. If two data show linear behavior, one can say that these sample data may come from the assumed distribution.

Sample Percentiles

Order the n sample observations from smallest to largest. Then the ith smallest observation in the list is taken to be the ***[100(i-0.5)/n]***th sample percentile.

$$\text{ith smallest sample} = [100(i-.5)/n] \text{ sample percentile}$$

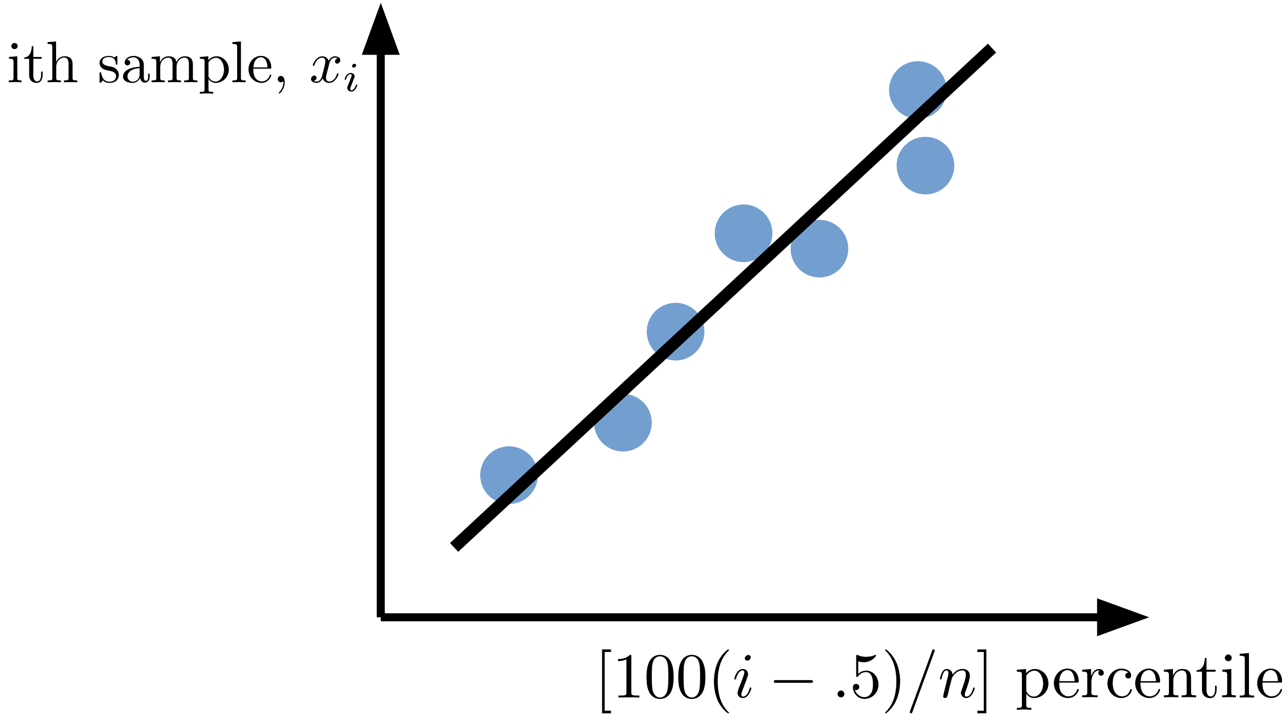
If the sample data comes from some specifying distribution,

$$\text{ith smallest sample} \approx [100(i-.5)/n] \text{ percentile of the distribution}$$

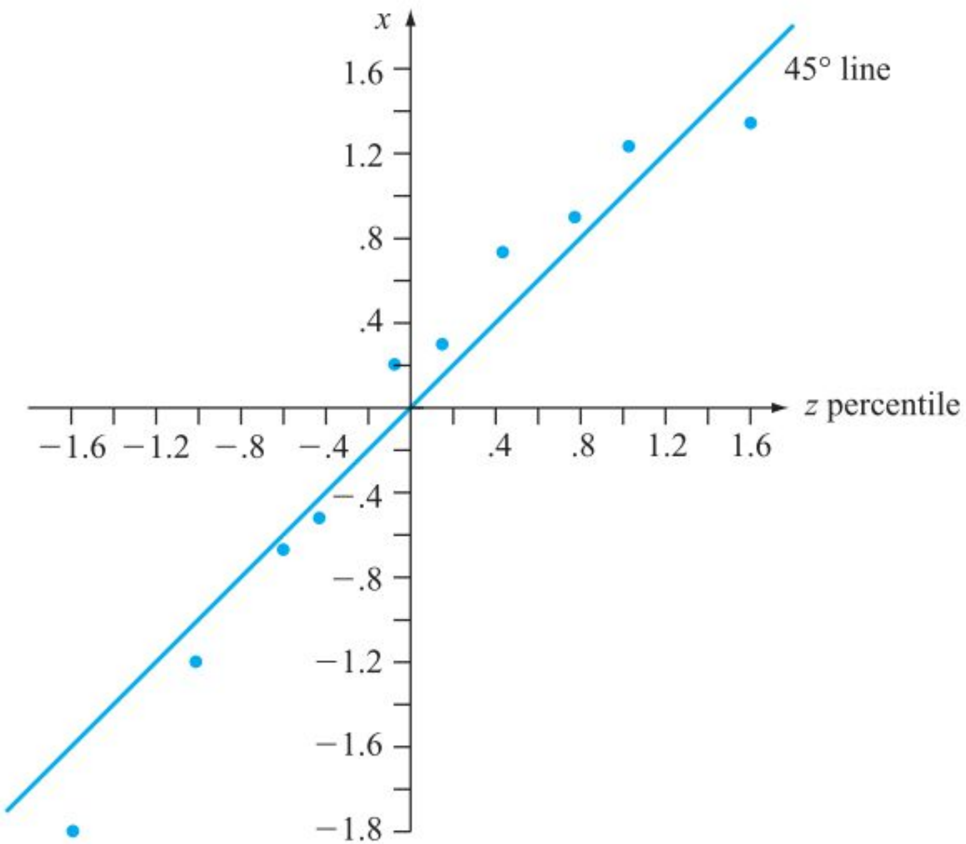
Probability Plot

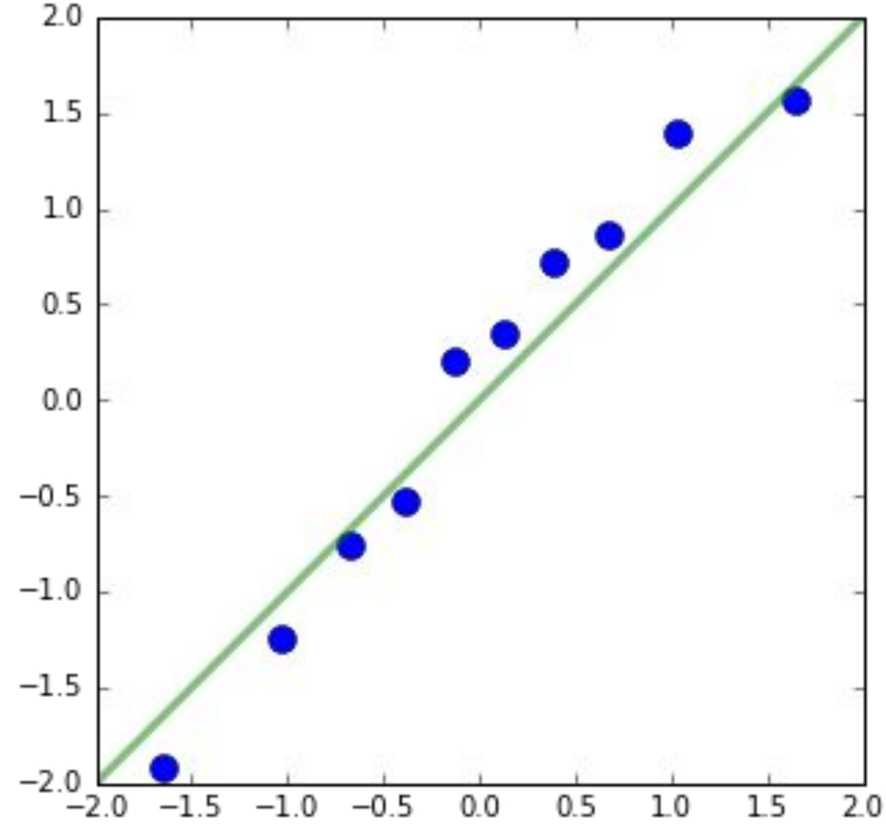
$$([100(i-.5)/n] \text{ percentile of the distribution}, \text{ith smallest sample})$$

If these pairs show a linear behavior, the sample data are derived from the assumed distribution of population.



Example: Probability Plot

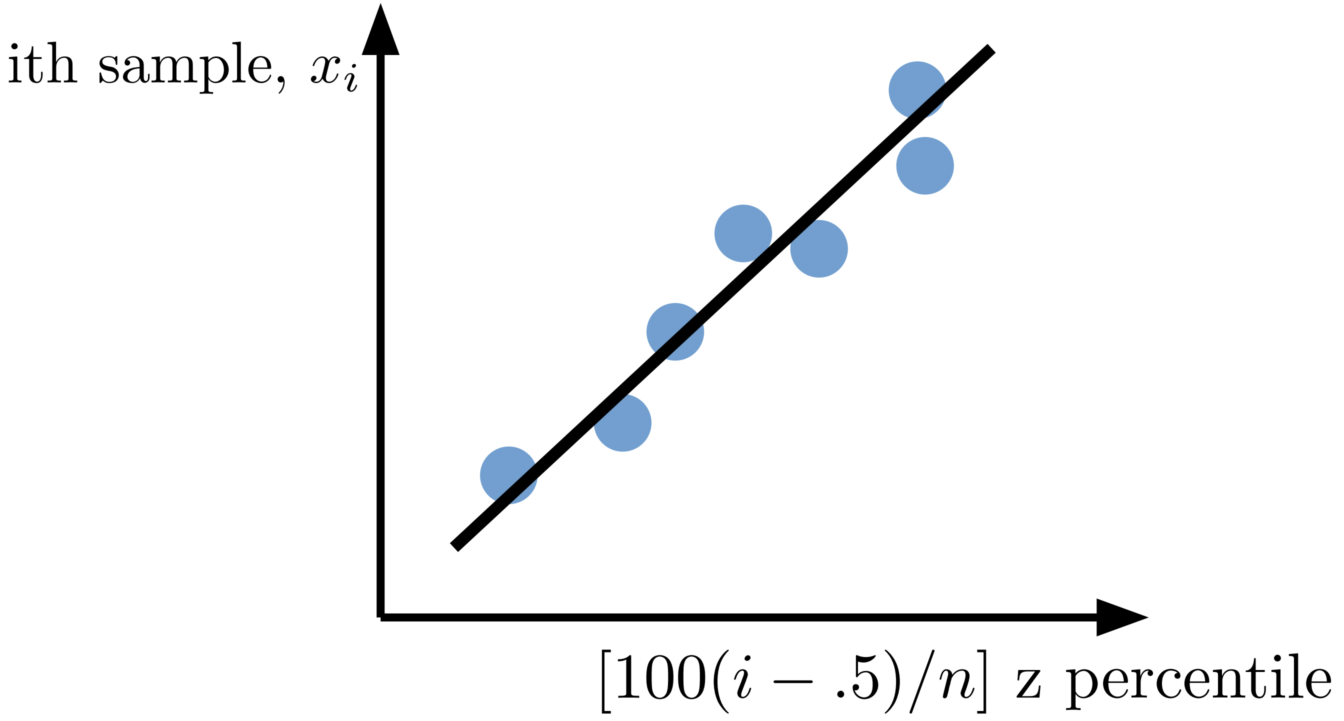




Normal Probability Plot

$$\left([100(i-.5)/n] \text{ percentile of } \underline{N(\mu,\sigma^2)}, \text{ith smallest sample}\right)$$

Normal distribution



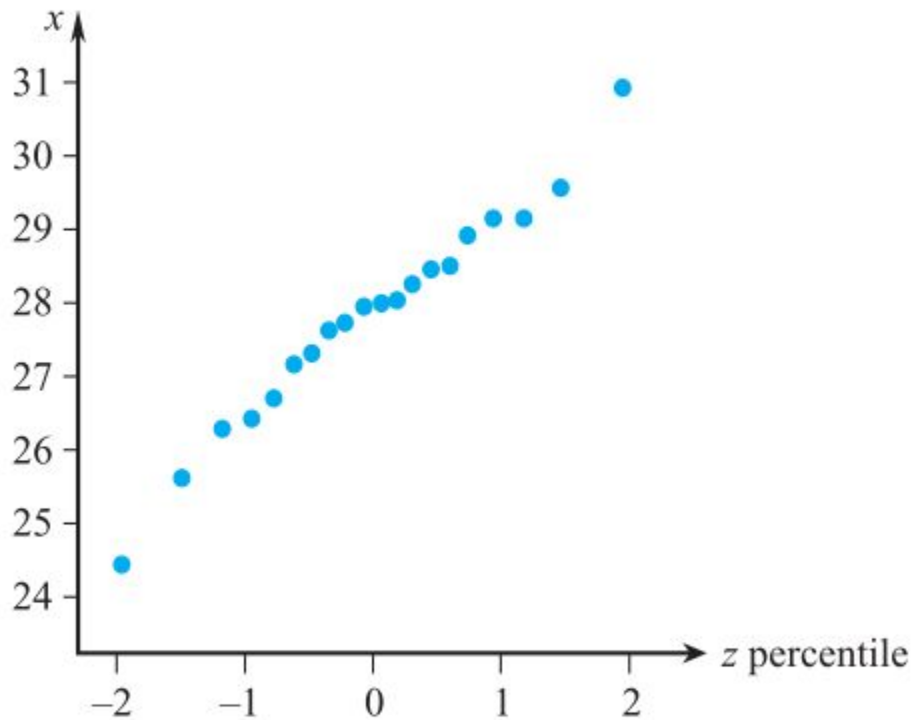
$$(100p)\text{th percentile for } N(\mu,\sigma^2) =$$

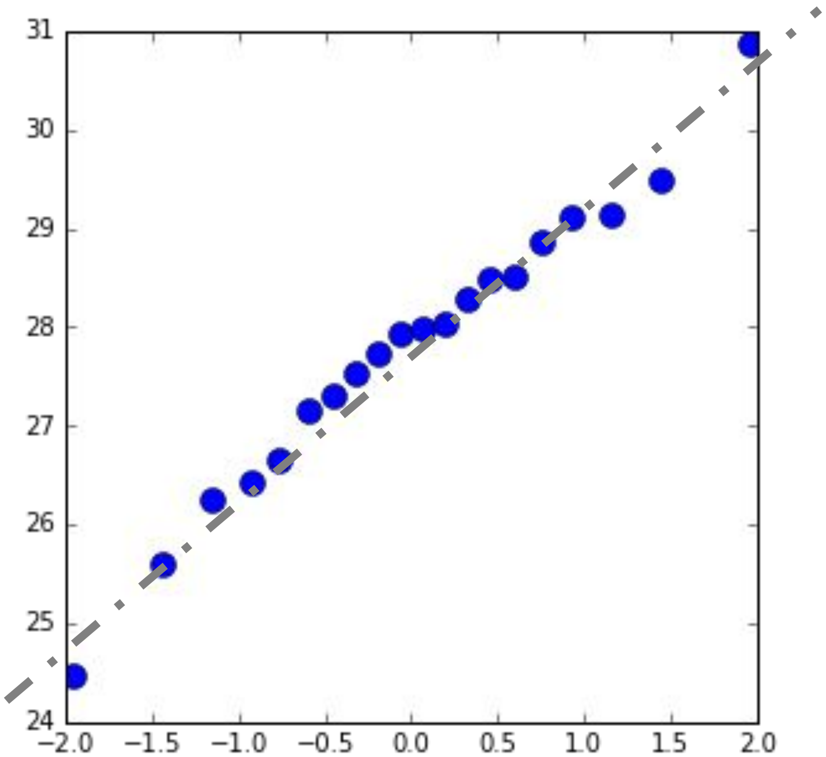
$$= \mu + [(100p)\text{th percentile for } N(0,1)]\,\sigma$$

Example: Normal Probability Plot

Observation	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
z percentile	−1.96	−1.44	−1.15	−.93	−.76	−.60	−.45	−.32	−.19	−.06
Observation	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88
z percentile	.06	.19	.32	.45	.60	.76	.93	1.15	1.44	1.96

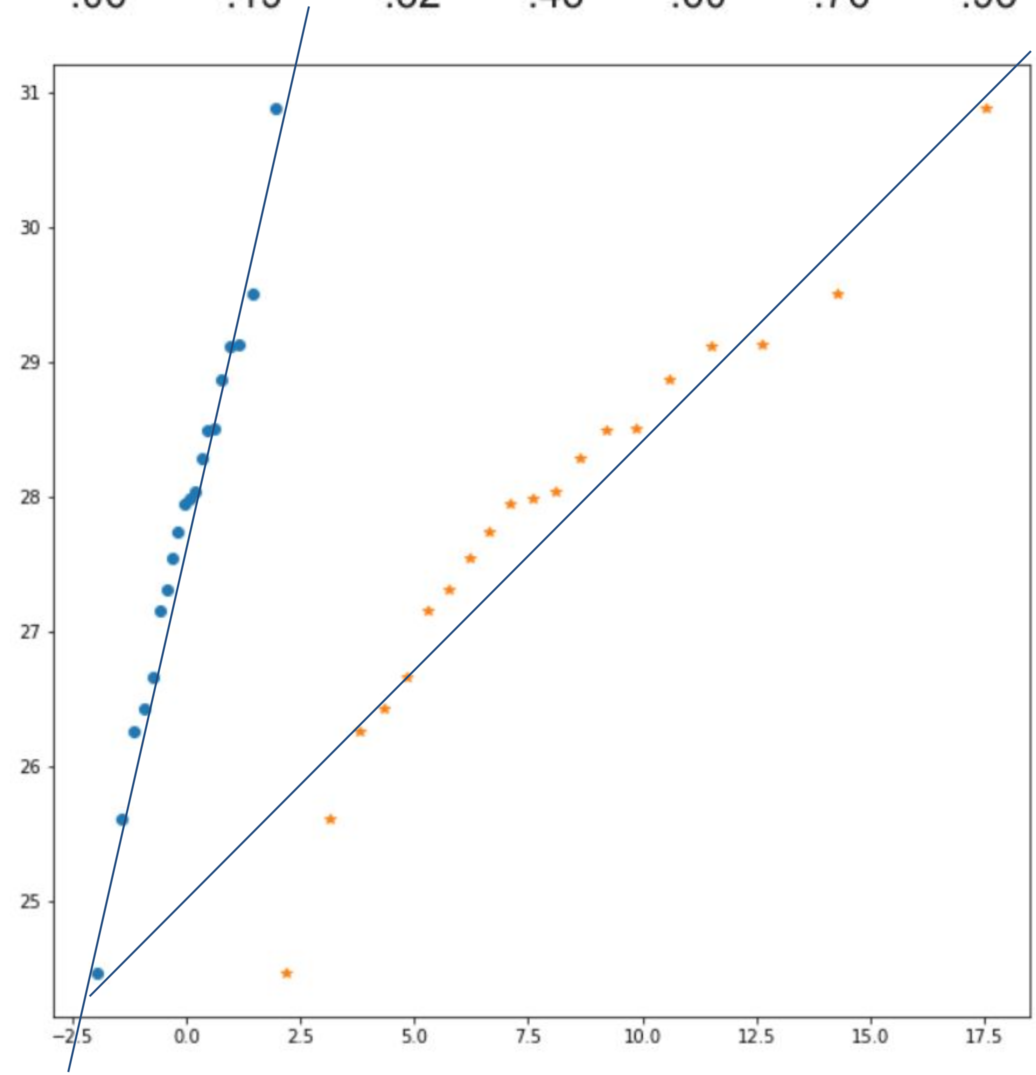
Probability and Statistics for Engineering and the Sciences, J. L. Devore



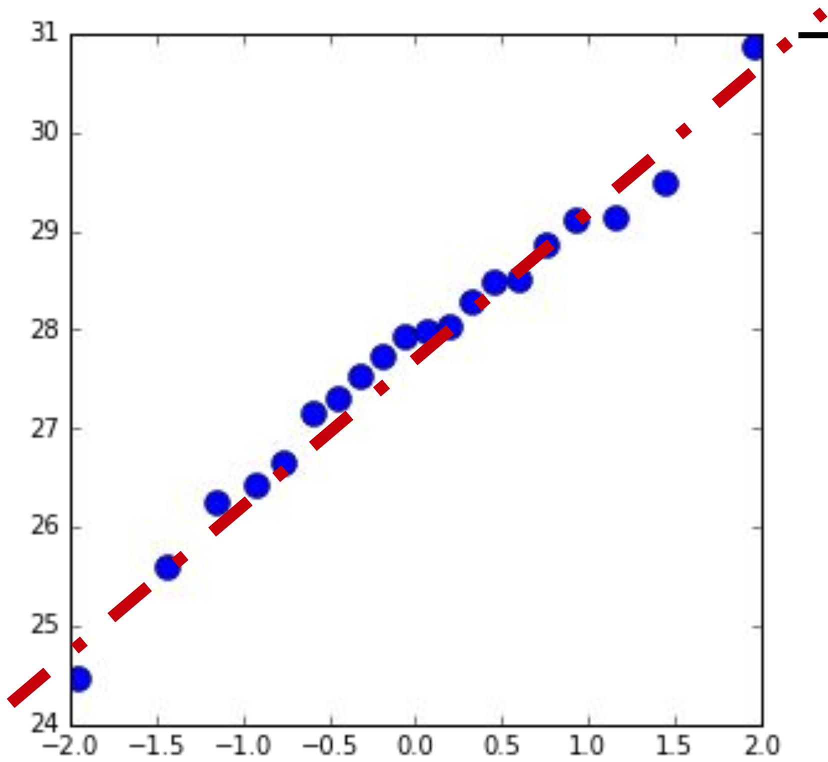


Example: Normal Probability Plot

Observation	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
z percentile	−1.96	−1.44	−1.15	−.93	−.76	−.60	−.45	−.32	−.19	−.06
Observation	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88
z percentile	.06	.19	.32	.45	.60	.76	.93	1.15	1.44	1.96



Example: Normal Probability Plot



How can we find out this linear equation?

Linear regression