50.034 - Introduction to Probability and Statistics

Week 5 – Cohort Class

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

CC Week 5 Assigned Readings: (Revise: 4.6, 4.7, 5.6)

Outline of Cohort Class

Exercises on the following topics:

- Covariance and correlation
- Conditional expectation
- Normal distributions

Recall: Covariance and Correlation

Let X and Y be R.V.'s with finite means μ_X and μ_Y respectively. The covariance of X and Y is $\mathrm{cov}(X, Y) = \mathbf{E}[(X - \mu_X)(Y - \mu_Y)]$, provided that this expectation $\mathbf{E}[(X - \mu_X)(Y - \mu_Y)]$ exists.

- $\mathrm{cov}(X, Y) =$ "how strongly X and Y are **linearly** related".
- **Useful Formula:** $\mathrm{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$
- **Note:** $\mathrm{var}(X) = \mathrm{cov}(X, X)$.

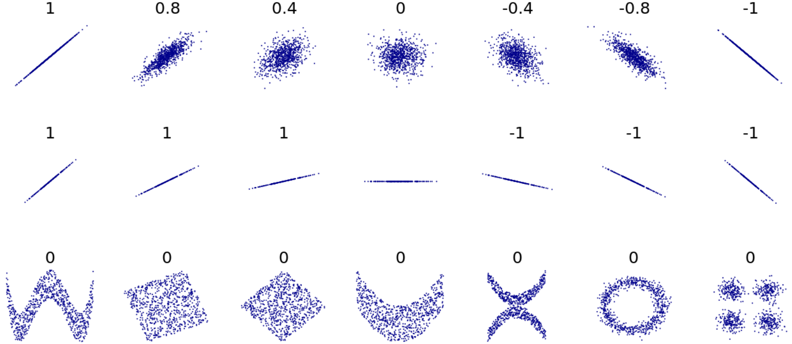
If X and Y have finite variances σ_X^2 and σ_Y^2 respectively, then the correlation of X and Y is

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

- **Fact:** $-1 \leq \rho(X, Y) \leq 1$ (by Cauchy–Schwarz inequality).
- Correlation does not depend on scale or choice of units.

Examples of different correlations

Let X and Y be the x-coordinate and y-coordinate of a randomly selected point on the xy-plane. Each plot below corresponds to 1000 points selected from some joint distribution.



Source: Denis Boigelot

Note: $Y = 0$ for the middle plot (i.e. $\mathrm{var}(Y) = 0$), so $\rho(X, Y)$ is undefined.

Last row: 7 different ways for X and Y to have no linear relations!

Exercise 1 (20 mins)

Let A, B, C be arbitrary R.V.'s with finite means and variances. Suppose $X = A + B$, $Y = B + C$ and $Z = C - A$.

1. Express $\mathrm{cov}(X, Y)$ in terms of the following:

$$\mathrm{cov}(A, B),\ \mathrm{cov}(A, C),\ \mathrm{cov}(B, C),\ \text{and } \mathrm{var}(B).$$

2. Express $\mathrm{cov}(Y, Z)$ in terms of the following:

$$\mathrm{cov}(A, B),\ \mathrm{cov}(A, C),\ \mathrm{cov}(B, C),\ \text{and } \mathrm{var}(C).$$

3. Express $\mathrm{cov}(X, Z)$ in terms of the following:

$$\mathrm{cov}(A, B),\ \mathrm{cov}(A, C),\ \mathrm{cov}(B, C),\ \text{and } \mathrm{var}(A).$$

Exercise 1 - Solution

1. [Expression for $\mathrm{cov}(X, Y) = \mathrm{cov}(A + B, B + C)$.]

Recall: (Lecture 5) "mean of sum" = "sum of means".
Hence, using the formula $\mathrm{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$,

$$\begin{aligned}
\mathrm{cov}(X, Y) &= \mathbf{E}[(A + B)(B + C)] - \mathbf{E}[A + B]\mathbf{E}[B + C] \\
&= \mathbf{E}[AB + AC + B^2 + BC] - (\mathbf{E}[A] + \mathbf{E}[B])(\mathbf{E}[B] + \mathbf{E}[C]) \\
&= \Big(\mathbf{E}[AB] + \mathbf{E}[AC] + \mathbf{E}[B^2] + \mathbf{E}[BC]\Big) \\
&\quad - \Big(\mathbf{E}[A]\mathbf{E}[B] + \mathbf{E}[A]\mathbf{E}[C] + \mathbf{E}[B]\mathbf{E}[B] + \mathbf{E}[B]\mathbf{E}[C]\Big) \\
&= (\mathbf{E}[AB] - \mathbf{E}[A]\mathbf{E}[B]) + (\mathbf{E}[AC] - \mathbf{E}[A]\mathbf{E}[C]) \\
&\quad + (\mathbf{E}[B^2] - \mathbf{E}[B]\mathbf{E}[B]) + (\mathbf{E}[BC] - \mathbf{E}[B]\mathbf{E}[C]) \\
&= \mathrm{cov}(A, B) + \mathrm{cov}(A, C) + \mathrm{var}(B) + \mathrm{cov}(B, C)
\end{aligned}$$

Exercise 1 - Solution

2. [Expression for $\operatorname{cov}(Y, Z) = \operatorname{cov}(B + C, C - A)$.]

Recall: (Lecture 5) "mean of sum" = "sum of means".
Hence, again using the covariance formula,

$$
\begin{aligned}
\operatorname{cov}(Y, Z) &= \mathbf{E}[(B + C)(C - A)] - \mathbf{E}[B + C]\mathbf{E}[C - A] \\
&= \mathbf{E}[BC - AB + C^2 - AC] - (\mathbf{E}[B] + \mathbf{E}[C])(\mathbf{E}[C] - \mathbf{E}[A]) \\
&= \Big(\mathbf{E}[BC] - \mathbf{E}[AB] + \mathbf{E}[C^2] - \mathbf{E}[AC]\Big) \\
&\quad - \Big(\mathbf{E}[B]\mathbf{E}[C] - \mathbf{E}[A]\mathbf{E}[B] + \mathbf{E}[C]\mathbf{E}[C] - \mathbf{E}[A]\mathbf{E}[C]\Big) \\
&= (\mathbf{E}[BC] - \mathbf{E}[B]\mathbf{E}[C]) - (\mathbf{E}[AB] - \mathbf{E}[A]\mathbf{E}[B]) \\
&\quad + (\mathbf{E}[C^2] - \mathbf{E}[C]\mathbf{E}[C]) - (\mathbf{E}[AC] - \mathbf{E}[A]\mathbf{E}[C]) \\
&= \operatorname{cov}(B, C) - \operatorname{cov}(A, B) + \operatorname{var}(C) - \operatorname{cov}(A, C)
\end{aligned}
$$

Exercise 1 - Solution

3. [Expression for $\mathrm{cov}(X, Z) = \mathrm{cov}(A + B, C - A)$.]

Recall: (Lecture 5) "mean of sum" = "sum of means".
Hence, again using the covariance formula,

$$\begin{aligned}
\mathrm{cov}(X, Z) &= \mathbf{E}[(A + B)(C - A)] - \mathbf{E}[A + B]\mathbf{E}[C - A] \\
&= \mathbf{E}[AC - A^2 + BC - AB] - (\mathbf{E}[A] + \mathbf{E}[B])(\mathbf{E}[C] - \mathbf{E}[A]) \\
&= \Big(\mathbf{E}[AC] - \mathbf{E}[A^2] + \mathbf{E}[BC] - \mathbf{E}[AB]\Big) \\
&\quad - \Big(\mathbf{E}[A]\mathbf{E}[C] - \mathbf{E}[A]\mathbf{E}[A] + \mathbf{E}[B]\mathbf{E}[C] - \mathbf{E}[A]\mathbf{E}[B]\Big) \\
&= (\mathbf{E}[AC] - \mathbf{E}[A]\mathbf{E}[C]) - (\mathbf{E}[A^2] - \mathbf{E}[A]\mathbf{E}[A]) \\
&\quad + (\mathbf{E}[BC] - \mathbf{E}[B]\mathbf{E}[C]) - (\mathbf{E}[AB] - \mathbf{E}[A]\mathbf{E}[B]) \\
&= \mathrm{cov}(A, C) - \mathrm{var}(A) + \mathrm{cov}(B, C) - \mathrm{cov}(A, B)
\end{aligned}$$

Covariance and variances

What we know: $\mathrm{var}(X) = \mathrm{cov}(X, X)$ and $\mathrm{var}(Y) = \mathrm{cov}(Y, Y)$.
Question: How are $\mathrm{cov}(X, X), \mathrm{cov}(Y, Y), \mathrm{cov}(X, Y)$ related?

Theorem: Let X and Y be R.V.'s with finite variances. Then for any constants a, b, c,

$$\mathrm{var}(aX + bY + c) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\,\mathrm{cov}(X, Y).$$

- Adding a constant to a R.V. does not change its variance.
- $\mathrm{var}(aX + bY + c) = \mathrm{var}(aX + bY)$.
- $\mathrm{var}(aX + bY) = a^2\mathrm{cov}(X, X) + b^2\mathrm{cov}(Y, Y) + 2ab\,\mathrm{cov}(X, Y)$.

Special cases:

- $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y)$.
- $\mathrm{var}(X - Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\mathrm{cov}(X, Y)$.
- If X and Y are independent, then $\mathrm{cov}(X, Y) = 0$, hence

  $$\mathrm{var}(aX + bY + c) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y)$$

  in this independent R.V.'s case.

Independence and expectation of R.V.'s

Theorem: If X and Y are **independent** R.V.'s with finite means, then

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y].$$

Proof:
We know that $\mathrm{cov}(X, Y) = 0$, since X and Y are independent.
We also know that $\mathrm{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$.
Therefore $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.

Note: Be careful! In general, $\mathbf{E}[X^2] \neq \mathbf{E}[X]\mathbf{E}[X]$.

- X is not independent of itself, except for the special case when $\Pr(X = x_0) = 1$ for some fixed real number x_0.

Theorem: If $X_1, \ldots, X_n$ are **independent** R.V.'s with finite means, then

$$\mathbf{E}[X_1 \cdots X_n] = \mathbf{E}[X_1] \cdots \mathbf{E}[X_n].$$

(See Theorem 4.2.6 in course textbook for a proof.)

More on independence and expectation

Theorem: Let $g_1(x), \ldots, g_n(x)$ be n real-valued functions, and let $X_1, \ldots, X_n$ be **independent** R.V.'s. Suppose that the expectations $\mathbf{E}[g(X_1)], \ldots, \mathbf{E}[g(X_n)]$ are all finite. Then

$$\mathbf{E}[g_1(X_1) \cdots g_n(X_n)] = \mathbf{E}[g_1(X_1)] \cdots \mathbf{E}[g_n(X_n)].$$

(Theorem 4.2.6 in the course textbook gives the case when every g_i is the identity function. Can you modify its proof to show this theorem?)

- In other words, "mean of product of functions" = "product of means of functions", under the assumption that the R.V.'s are **independent**.

Exercise 2 (15 mins)

Let X, Y, Z be R.V.'s with finite means and finite variances.

Suppose that X and Y are positively correlated, i.e. $\rho(X, Y) > 0$.
Suppose also Y and Z are positively correlated, i.e. $\rho(Y, Z) > 0$.

Is it true that X and Z must necessarily be positively correlated?

Hint: In Exercise 1, what if A, B, C are independent R.V.'s?
Can you think of a "real-world" example?

Exercise 2 - Solution

In Exercise 1, we showed that if A, B, C are R.V.'s with finite means and variances, and if $X = A + B$, $Y = B + C$, $Z = C - A$, then

$$\begin{aligned}
\mathrm{cov}(X, Y) &= \mathrm{cov}(A, B) + \mathrm{cov}(A, C) + \mathrm{var}(B) + \mathrm{cov}(B, C); \\
\mathrm{cov}(Y, Z) &= \mathrm{cov}(B, C) - \mathrm{cov}(A, B) + \mathrm{var}(C) - \mathrm{cov}(A, C); \\
\mathrm{cov}(X, Z) &= \mathrm{cov}(A, C) - \mathrm{var}(A) + \mathrm{cov}(B, C) - \mathrm{cov}(A, B).
\end{aligned}$$

If A, B, C are **independent** R.V.'s, then the covariances simplify to:

$$\begin{aligned}
\mathrm{cov}(X, Y) &= \mathrm{var}(B); \\
\mathrm{cov}(Y, Z) &= \mathrm{var}(C); \\
\mathrm{cov}(X, Z) &= -\mathrm{var}(A).
\end{aligned}$$

Exercise 2 - Solution

Suppose further that A, B, C all have strictly positive variances.

$$\text{var}(X) = \text{var}(A + B) = \text{var}(A) + \text{var}(B) > 0;$$
$$\text{var}(Y) = \text{var}(B + C) = \text{var}(B) + \text{var}(C) > 0;$$
$$\text{var}(Z) = \text{var}(C - A) = \text{var}(C) + \text{var}(A) > 0.$$

Consequently,

$$\text{cov}(X, Y) = \text{var}(B) > 0 \Rightarrow \rho(X, Y) = \frac{cov(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} > 0;$$
$$\text{cov}(Y, Z) = \text{var}(C) > 0 \Rightarrow \rho(Y, Z) = \frac{cov(Y,Z)}{\sqrt{\text{var}(Y)}\sqrt{\text{var}(Z)}} > 0;$$
$$\text{cov}(X, Z) = -\text{var}(A) < 0 \Rightarrow \rho(X, Z) = \frac{cov(X,Z)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Z)}} < 0.$$

In other words, we can always find R.V.'s X, Y, Z such that:

- X and Y are positively correlated;
- Y and Z are positively correlated;
- BUT X and Z are **negatively** correlated!

"Real-world" Example for Exercise 2

In basketball, NBA basketball players could play as many as > 80 games per year, and their points scored over the year are recorded. For a randomly selected NBA basketball player:

- Let X be the number of slam dunks made in a year.
- Let Y be the total number of points scored in a year.
- Let Z be the number of 3-pointers made in a year.

1. X and Y are positively correlated.
   - More slam dunks $\Rightarrow$ Greater tendency to score more points.
2. Y and Z are positively correlated.
   - More points $\Rightarrow$ Greater tendency to make more 3-pointers.
3. However, X and Z are **negatively** correlated!
   - More slam dunks $\Rightarrow$ Greater tendency to make **less** 3-pointers.

This makes sense, since it is quite hard to miss a slam dunk.

- A player who is good at slam dunks would score more from slam dunks and tend to go closer to the basketball hoop to score, instead of trying to score from 3-pointers, which has a success rate of $\approx 40\%$ even for the best 3-point shooters.

Conditional expectation

Recall: Conditional expectation $=$ $\mathbf{E}$[conditional distribution].

Let X, Y be R.V.'s, and let $C' \subseteq \mathbb{R}$ such that $\Pr(Y \in C') > 0$.
Definition: The conditional expectation of X given $Y \in C'$ is:

$$\mathbf{E}[X|Y \in C'] = \mathbf{E}\left[\begin{array}{c}\text{conditional distribution} \\ \text{of } X \text{ given } Y \in C'.\end{array}\right]$$

- If C' is fixed, then $\mathbf{E}[X|Y \in C']$ is a fixed value.

If we are given a specific value $Y = y$, then the conditional expectation of X given $Y = y$ is denoted by $\mathbf{E}[X|Y = y]$.

- Similarly, if y is fixed, then $\mathbf{E}[X|Y = y]$ is a fixed value.

Conditional expectation as a variable:

- We can think of $\mathbf{E}[X|Y \in C']$ as a function in terms of C'.
  - Different values of C' give different values for $\mathbf{E}[X|Y \in C']$.
- Similarly, we can think of $\mathbf{E}[X|Y = y]$ as a function of y.
  - Different values of y give different values for $\mathbf{E}[X|Y = y]$.

More generally, we can think of $\mathbf{E}[X|Y]$ as a **function of** Y.

- In other words, $\mathbf{E}[X|Y]$ is a random variable!

Exercise 3 (15 mins)

An experiment involves rolling a fair die as many times as necessary until the first 6 is rolled.

Let X be the total number of 1's we get in the experiment.
Let Y be the number of rolls made before the first 6 is rolled.

1. What is $\mathbf{E}[X|Y = y]$ in terms of y?
2. What does $\mathbf{E}[\mathbf{E}[X|Y]]$ mean, and what is its value?
3. What is the value of $\mathbf{E}[X]$?

Exercise 3 - Solution

1. [What is $\mathbf{E}[X|Y = y]$ in terms of y?]

If $Y = y$, then the first y rolls must not contain any 6's, and the $(y + 1)$-th roll must be a 6.
Among the first y rolls, each number $1, 2, 3, 4, 5$ is equally likely to be rolled, so each of the y rolls has a $\frac{1}{5}$ probability of being a 1.
Therefore, $\mathbf{E}[X|Y = y] = \frac{y}{5}$.
Note: If y is a fixed value, then $\mathbf{E}[X|Y = y]$ is a value.

- For example, $\mathbf{E}[X|Y = 10] = \frac{10}{5} = 2$.

2. [What does $\mathbf{E}[\mathbf{E}[X|Y]]$ mean, and what is its value?]

$\mathbf{E}[X|Y]$ is a R.V. representing the number X of 1's rolled, given the total number Y of rolls in the experiment.
Hence $\mathbf{E}[\mathbf{E}[X|Y]]$ is the expectation of this R.V., or equivalently, the expected number of 1's rolled, given the total number of rolls in the experiment.
From part 1., $\mathbf{E}[X|Y]$ is precisely the R.V. $\frac{Y}{5}$.

Exercise 3 - Solution

2. (continued)
Note that Y can be modeled as a geometric R.V. with parameter $p = \frac{1}{6}$, so its pmf is

$$p(y) = \begin{cases} \frac{1}{6} \cdot (\frac{5}{6})^y, & \text{if } y = 0, 1, 2, \ldots; \\ 0, & \text{otherwise.} \end{cases}$$

From Lecture 6, a geometric R.V. with parameter p has mean $\frac{1-p}{p}$, hence $\mathbf{E}[Y] = (1 - \frac{1}{6})/\frac{1}{6} = 5$, which implies

$$\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[\tfrac{1}{5}Y] = \tfrac{1}{5} \cdot 5 = 1.$$

3. [What is the value of $\mathbf{E}[X]$?]

By the law of total probability for expectations, $\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X]$, therefore from part 2., $\mathbf{E}[X] = 1$.

From normal R.V. to standard normal R.V.

Let $X \sim N(\mu, \sigma^2)$, and let $F(x)$ be the cdf of X.
Let $Z \sim N(0,1)$, and let $\Phi(z)$ be the cdf of Z.
Fact: $\frac{X-\mu}{\sigma}$ and Z have the exact same distributions.

- Note: The denominator is standard deviation, not variance!

How to calculate $\Pr(\mathbf{a} \le \mathbf{X} \le \mathbf{b})$?

$$\Pr(a \le X \le b) = \Pr\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Similarly,

$$\Pr(X \ge a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$
$$\Pr(X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Useful Fact: $\Phi(-z) = 1 - \Phi(z)$ for all real numbers z.

Example 1

The error in distance measurement by a GPS system follows a normal distribution with a mean of 5 cm and a standard deviation of 3 cm.

What is the probability that the measurement error is between 4.1 cm and 6.8 cm?

Some potentially useful values:

$$\Phi(0.10) \approx 0.5398,\ \Phi(0.20) \approx 0.5793,\ \Phi(0.30) \approx 0.6179,$$
$$\Phi(0.40) \approx 0.6554,\ \Phi(0.50) \approx 0.6915,\ \Phi(0.60) \approx 0.7257.$$

($\Phi(z)$ denotes the standard normal cdf.)

Example 1 - Solution

Let X be the normal R.V. representing this distance error (in cm). We are given that X has mean $\mu = 5$ and standard deviation $\sigma = 3$.

Define the R.V. $Z = \frac{X-\mu}{\sigma} = \frac{X-5}{3}$. Then Z has the standard normal distribution.

Therefore, the probability is

$$\begin{aligned}
\Pr(4.1 \leq X \leq 6.8) &= \Pr\left(\frac{4.1-5}{3} \leq Z \leq \frac{6.8-5}{3}\right) \\
&= \Pr(-0.3 \leq Z \leq 0.6) \\
&= \Pr(Z \leq 0.6) - \Pr(Z \leq -0.3) \\
&= \Phi(0.6) - \Phi(-0.3) \\
&= \Phi(0.6) - (1 - \Phi(0.3)) \\
&\approx 0.7257 + 0.6179 - 1 \\
&= 0.3436
\end{aligned}$$

Exercise (10 mins)

The amount of waste gas emitted by a randomly selected vehicle can be modeled by a normal distribution with a mean of 7000 ppl (particles per litre) and a standard deviation of 1300 ppl.

We want to set a threshold for the maximum acceptable emission level. What shoud this threshold be, so that 10% of the vehicles will fail to pass the emission test?

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932
0.08	0.5319	0.68	0.7517	1.28	0.8997	1.88	0.9699	2.48	0.9934
0.09	0.5359	0.69	0.7549	1.29	0.9015	1.89	0.9706	2.49	0.9936
0.10	0.5398	0.70	0.7580	1.30	0.9032	1.90	0.9713	2.50	0.9938
0.11	0.5438	0.71	0.7611	1.31	0.9049	1.91	0.9719	2.52	0.9941
0.12	0.5478	0.72	0.7642	1.32	0.9066	1.92	0.9726	2.54	0.9945
0.13	0.5517	0.73	0.7673	1.33	0.9082	1.93	0.9732	2.56	0.9948
0.14	0.5557	0.74	0.7704	1.34	0.9099	1.94	0.9738	2.58	0.9951
0.15	0.5596	0.75	0.7734	1.35	0.9115	1.95	0.9744	2.60	0.9953
0.16	0.5636	0.76	0.7764	1.36	0.9131	1.96	0.9750	2.62	0.9956
0.17	0.5675	0.77	0.7794	1.37	0.9147	1.97	0.9756	2.64	0.9959
0.18	0.5714	0.78	0.7823	1.38	0.9162	1.98	0.9761	2.66	0.9961

Exercise - Solution

Let X be the normal R.V. representing the emission level (in ppl), and let the threshold value (in ppl) be t.

"10% of the vehicles fail the emission test" is the same as "10% of the vehicles emit waste gas whose amount exceeds the threshold".

- i.e. $\Pr(X \geq t) = 0.1$, or equivalently, $\Pr(X \leq t) = 0.9$.

Define $Z = \frac{X-\mu}{\sigma}$, and note that $Z \sim N(0,1)$. Then we have:

$$\Pr(X \leq t) = \Pr\left(Z \leq \frac{t-\mu}{\sigma}\right) = \Phi(\tfrac{t-\mu}{\sigma}) = 0.9.$$

Checking the standard normal distribution table, the closest value we can find for z satisfying $\Phi(z) = 0.9$ is $z = 1.28$.

- In other words, $z = 1.28$ is approximately the 90-th percentile.

Thus, $\frac{t-\mu}{\sigma} = \frac{t-7000}{1300} \approx 1.28$, which implies:

$$t \approx (1.28)(1300) + 7000 = 8664.$$

Therefore, the threshold should be approximately 8664 ppl.

Exercise (10 mins)

The breakdown voltage of a randomly chosen LED of particular type is known to follow a normal distribution with a mean of 20 volts, and a standard deviation of 1.2 volts.

What is the probability that a LED's breakdown voltage is within two standard deviations of its mean?

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932
0.08	0.5319	0.68	0.7517	1.28	0.8997	1.88	0.9699	2.48	0.9934
0.09	0.5359	0.69	0.7549	1.29	0.9015	1.89	0.9706	2.49	0.9936
0.10	0.5398	0.70	0.7580	1.30	0.9032	1.90	0.9713	2.50	0.9938
0.11	0.5438	0.71	0.7611	1.31	0.9049	1.91	0.9719	2.52	0.9941
0.12	0.5478	0.72	0.7642	1.32	0.9066	1.92	0.9726	2.54	0.9945
0.13	0.5517	0.73	0.7673	1.33	0.9082	1.93	0.9732	2.56	0.9948
0.14	0.5557	0.74	0.7704	1.34	0.9099	1.94	0.9738	2.58	0.9951
0.15	0.5596	0.75	0.7734	1.35	0.9115	1.95	0.9744	2.60	0.9953
0.16	0.5636	0.76	0.7764	1.36	0.9131	1.96	0.9750	2.62	0.9956
0.17	0.5675	0.77	0.7794	1.37	0.9147	1.97	0.9756	2.64	0.9959
0.18	0.5714	0.78	0.7823	1.38	0.9162	1.98	0.9761	2.66	0.9961
0.19	0.5753	0.79	0.7852	1.39	0.9177	1.99	0.9767	2.68	0.9963
0.20	0.5793	0.80	0.7881	1.40	0.9192	2.00	0.9773	2.70	0.9965

Exercise - Solution

Let X be the normal R.V. representing the breakdown voltage (in volts). X has mean $\mu = 20$ and standard deviation $\sigma = 1.2$.
We want to find the probability $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$.
Let $Z = \frac{X-\mu}{\sigma}$, and note that $Z \sim N(0,1)$. Thus,

$$\begin{aligned}
\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= \Pr\left(\frac{\mu - 2\sigma - \mu}{\sigma} \leq Z \leq \frac{\mu + 2\sigma - \mu}{\sigma}\right) \\
&= \Pr(-2 \leq Z \leq 2) \\
&= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \\
&\approx 0.9546
\end{aligned}$$

Note: We do not need to know the values of μ or σ.

- We only used the condition that X has a normal distribution.

Consequence: Given **any normal R.V.**, the probability that the R.V. is within two standard deviations will always be ≈ 0.9546.

- Similarly, $\Pr(\mu - \sigma \leq X \leq \mu + \sigma) = 2\Phi(1) - 1 \approx 0.6826$, which does not depend on the exact values of μ and σ.

Summary

Exercises on the following topics:

- Covariance and correlation
- Conditional expectation
- Normal distributions

Reminders:

- There is **mini-quiz 2** next Cohort Class!
  - Tested on all materials from Lectures 6–10 and Cohort classes weeks 3–5. (Today's class is Week 5 Cohort Class.)
- Homework Set 5 is also due next Cohort Class.