Set 30 Let $X_1, \dots, X_n$ be independent and identically distributed random variables with the same mean μ and variance σ^2, and let $\bar{X}$ denote the sample mean. Tchebychev's inequality says that the probability $\mathbb{P}(|\bar{X}-\mu| > \delta)$ does not exceed $\mathrm{Var}(\bar{X})/\delta^2$ for every constant $\delta > 0$, where $\mathrm{Var}(\bar{X})$ denotes the variance of $\bar{X}$.

(a) Prove that the statement

$$|\bar{X}-\mu| > \sqrt{20}\,\frac{\sigma}{\sqrt{n}}$$

holds with a probability that does not exceed 0.05.

Then,

$$\mathbb{P}(|\bar{X}-\mu| > \delta) \le \frac{\mathrm{Var}(\bar{X})}{\delta^2}$$

$$\mathbb{P}\left(|\bar{X}-\mu| > \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right) \le \frac{\mathrm{Var}(\bar{X})}{\left(\sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right)^2}$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Finally,

$$\mathbb{P}\left(|\bar{X}-\mu| > \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right) \le \frac{\frac{\sigma^2}{n}}{20\,\frac{\sigma^2}{n}}$$

$$\mathbb{P}\left(|\bar{X}-\mu| > \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right) \le \frac{1}{20}$$

(b) Prove that the statement

$$\bar{X}-\sqrt{20}\,\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}+\sqrt{20}\,\frac{\sigma}{\sqrt{n}}$$

holds with a probability that is not smaller than 0.95. NOTE: the previous statement gives the following confidence interval

$$\left[\bar{X}-\sqrt{20}\,\frac{\sigma}{\sqrt{n}}, \bar{X}+\sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right]$$

for the (unknown) population mean μ, with a confidence level of at least 95% and for every sample size n.

Then,

$$\bar{X} - \sqrt{20}\ \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \sqrt{20}\ \frac{\sigma}{\sqrt{n}}$$

$$-\bar{X} + \sqrt{20}\ \frac{\sigma}{\sqrt{n}} \geq -\mu \geq -\bar{X} - \sqrt{20}\ \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{20}\ \frac{\sigma}{\sqrt{n}} \geq \bar{X} - \mu \geq -\sqrt{20}\ \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{20}\ \frac{\sigma}{\sqrt{n}} \geq |\bar{X} - \mu|$$

$$|\bar{X} - \mu| \leq \sqrt{20}\ \frac{\sigma}{\sqrt{n}}$$

However,

$$\mathbb{P}(|\bar{X} - \mu| > \delta) \leq \frac{\mathrm{Var}(\bar{X})}{\delta^2}$$

But,

$$\mathbb{P}(|\bar{X} - \mu| \leq \delta) \leq 1 - \frac{\mathrm{Var}(\bar{X})}{\delta^2}$$

Therefore,

$$\mathbb{P}\left(|\bar{X} - \mu| \leq \sqrt{20}\ \frac{\sigma}{\sqrt{n}}\right) \leq 1 - \frac{\frac{\sigma^2}{n}}{\left(\sqrt{20}\ \frac{\sigma}{\sqrt{n}}\right)^2}$$

$$\mathbb{P}\left(|\bar{X} - \mu| \leq \sqrt{20}\ \frac{\sigma}{\sqrt{n}}\right) \leq 1 - \frac{\frac{\sigma^2}{n}}{20\ \frac{\sigma^2}{n}}$$

$$\mathbb{P}\left(|\bar{X} - \mu| \leq \sqrt{20}\ \frac{\sigma}{\sqrt{n}}\right) \leq 1 - \frac{1}{20}$$

$$\mathbb{P}\left(|\bar{X} - \mu| \leq \sqrt{20}\ \frac{\sigma}{\sqrt{n}}\right) \leq \frac{19}{20}$$

(c) Prove that the statement

$$\mu - \sqrt{20}\,\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + \sqrt{20}\,\frac{\sigma}{\sqrt{n}}$$

holds with a probability that is not smaller than 0.95.
Then,

$$\mu - \sqrt{20}\,\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + \sqrt{20}\,\frac{\sigma}{\sqrt{n}}$$

$$-\sqrt{20}\,\frac{\sigma}{\sqrt{n}} \le \bar{X} - \mu \le \sqrt{20}\,\frac{\sigma}{\sqrt{n}}$$

$$|\bar{X} - \mu| \le \sqrt{20}\,\frac{\sigma}{\sqrt{n}}$$

And,

$$\mathbb{P}(|\bar{X} - \mu| \le \delta) \le 1 - \frac{\mathrm{Var}(\bar{X})}{\delta^2}$$

Therefore,

$$\mathbb{P}\left(|\bar{X} - \mu| \le \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right) \le 1 - \frac{\frac{\sigma^2}{n}}{\left(\sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right)^2}$$

$$\mathbb{P}\left(|\bar{X} - \mu| \le \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right) \le 1 - \frac{\frac{\sigma^2}{n}}{20\,\frac{\sigma^2}{n}}$$

$$\mathbb{P}\left(|\bar{X} - \mu| \le \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right) \le 1 - \frac{1}{20}$$

$$\mathbb{P}\left(|\bar{X} - \mu| \le \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right) \le \frac{19}{20}$$

(d) Let $\omega_1{}^{act}, \dots, \omega_n{}^{act}$ be a sample from a population Ω, and let $X: \Omega \to \mathrm{R}$ be a filter (that is, a random variable in Statistics, or a measurable function in Mathematics) that produces n outputs (called observations in Statistics) $x_1{}^{obs}, \dots, x_n{}^{obs}$ by the formula $x_i{}^{obs} = X(\omega_i{}^{act})$. Let $\bar{x}$ denote the average of $x_1{}^{obs}, \dots, x_n{}^{obs}$. Does the confidence interval

$$\left[\bar{x} - \sqrt{20}\,\frac{\sigma}{\sqrt{n}}, \bar{x} + \sqrt{20}\,\frac{\sigma}{\sqrt{n}}\right]$$

cover the (unknown) population mean μ or not?

The confidence interval covers the population mean with a confidence percentage of 95%.

Set 31 Let $X_1, \dots, X_n$ be independent and identically distributed random variables with the same mean μ and variance σ^2, and let $\bar{X}$ denote the sample mean. Use the Central Limit Theorem (CLT, page 156 in our textbox) to establish the following (asymptotic when $n \to \infty$) statements:

(a) The statement

$$|\bar{X} - \mu| > 1.96 \frac{\sigma}{\sqrt{n}}$$

holds with probability 0.05.

Using the CLT, as $n \to \infty$

$$|G| = \frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}}$$

Therefore,

$$\mathbb{P}(|G| > 1.96) = 1 - \mathbb{P}(|G| \leq 1.96)$$

But since,

$$\mathbb{P}(|G| \leq 1.96) = 0.95$$

Then,

$$1 - \mathbb{P}(|G| \leq 1.96) = 1 - 0.95$$

$$\mathbb{P}(|G| > 1.96) = 0.05$$

(b) The statement

$$\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}$$

holds with the probability 0.95. NOTE: The statement about give the following (asymptotic when $n \to \infty$) 95% confidence interval

$$\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

Then,

$$\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}$$

$$-\bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \geq -\mu \geq -\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}$$

$$1.96 \frac{\sigma}{\sqrt{n}} \geq \bar{X} - \mu \geq -1.96 \frac{\sigma}{\sqrt{n}}$$

$$|\bar{X} - \mu| \leq 1.96 \frac{\sigma}{\sqrt{n}}$$

Therefore, using the CLT, as $n \to \infty$

$$|G| = \frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}}$$

And,

$$\mathbb{P}(|G| \leq 1.96) = 0.95$$

(c) The statement

$$\mu - 1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \frac{\sigma}{\sqrt{n}}$$

holds with the probability 0.95.

Then,

$$\mu - 1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \frac{\sigma}{\sqrt{n}}$$

$$-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}}$$

$$|\bar{X} - \mu| \leq 1.96 \frac{\sigma}{\sqrt{n}}$$

Therefore, using the CLT, as $n \to \infty$

$$|G| = \frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}}$$

And,

$$\mathbb{P}(|G| \leq 1.96) = 0.95$$

(d) Let $\omega_1{}^{act}, \dots, \omega_n{}^{act}$ be a sample from a population Ω, and let $X\colon \Omega \to \mathrm{R}$ be a filter (that is, a random variable in Statistics, or a measurable function in Mathematics) that produces n outputs (called observations in Statistics) $x_1{}^{obs}, \dots, x_n{}^{obs}$ by the formula $x_i{}^{obs} = X(\omega_i{}^{act})$. Let $\bar{x}$ denote the average of $x_1{}^{obs}, \dots, x_n{}^{obs}$. Does the 95% confidence interval

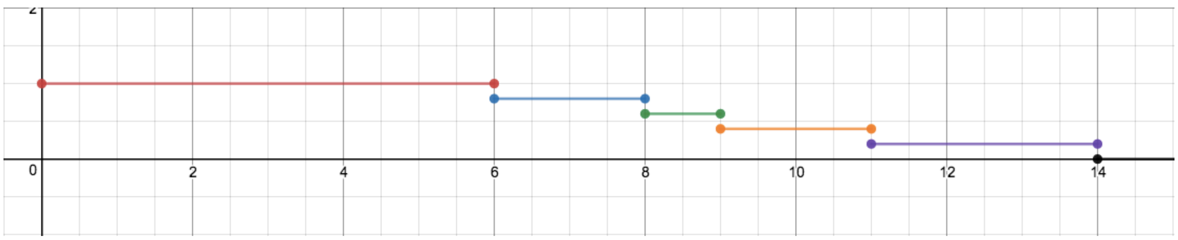
$$\left[\bar{x} - 1.96\,\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\,\frac{\sigma}{\sqrt{n}}\right]$$

cover the (unknown) population mean μ or not?

The confidence interval covers the population mean with a confidence percentage of 95%.

Set 32 Suppose that you own a wind farm with five wind turbines of the same type, and suppose that you have kept the records of their first failure times ${t_1}^{obs}, \ldots, {t_5}^{obs}$, which are 14, 6, 8, 11, and 9 months.

(a) Draw the sample (empirical) reliability function.



(b) Calculate the area under the empirical reliability function.

$$A = 6 * 1 + 2 * \frac{4}{5} + 1 * \frac{3}{5} + 2 * \frac{2}{5} + 3 * \frac{1}{5}$$

$$A = 9 + \frac{3}{5} = 9.6$$

(c) Calculate the sample mean $\bar{t}$ of the data.

$$\bar{t} = \frac{(6 + 8 + 9 + 11 + 14)}{5} = 9.6$$

(d) Construct a 95% (asymptotic) confidence interval for the population MTTF (mean time to failure) when someone has told you that, based on historical records, the population variance σ^2 is 9.

$$\bar{t} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{t} + 1.96\frac{\sigma}{\sqrt{n}}$$

$$9.6 - 1.96\frac{\sqrt{9}}{\sqrt{5}} \leq \mu \leq 9.6 + 1.96\frac{\sqrt{9}}{\sqrt{5}}$$

$$9.6 - 1.96\frac{3}{\sqrt{5}} \leq \mu \leq 9.6 + 1.96\frac{3}{\sqrt{5}}$$

Therefore, the 95% confidence interval is

$$\left[9.6 - 1.96\frac{3}{\sqrt{5}}, 9.6 + 1.96\frac{3}{\sqrt{5}}\right]$$

(e) Calculate the sample standard deviation $\hat{\sigma}$ of the data.

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - \bar{t})}$$

$$\hat{\sigma} = \sqrt{\frac{1}{5}((6-9.6)^2 + (8-9.6)^2 + (9-9.6)^2 + (11-9.6)^2 + (14-9.6)^2)}$$

$$\hat{\sigma} = \sqrt{\frac{1}{5}(3.6^2 + 1.6^2 + 0.6^2 + 1.4^2 + 4.4^2)}$$

$$\hat{\sigma} = \sqrt{\frac{1}{5}(12.96 + 2.56 + 0.36 + 1.96 + 19.36)}$$

$$\hat{\sigma} = \sqrt{\frac{37.2}{5}}$$

(f) Construct a 95% (asymptotic) confidence interval for the population MTTF when no information about the population variance σ^2 is available.

$$\bar{t} - 1.96\frac{\hat{\sigma}}{\sqrt{n}} \le \mu \le \bar{t} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}$$

$$9.6 - 1.96\frac{\sqrt{\frac{37.2}{5}}}{\sqrt{5}} \le \mu \le 9.6 + 1.96\frac{\sqrt{\frac{37.2}{5}}}{\sqrt{5}}$$

$$9.6 - 1.96\frac{\sqrt{37.2}}{\sqrt{5}}\frac{1}{\sqrt{5}} \le \mu \le 9.6 + 1.96\frac{\sqrt{37.2}}{\sqrt{5}}\frac{1}{\sqrt{5}}$$

$$9.6 - 1.96\frac{\sqrt{37.2}}{5} \le \mu \le 9.6 + 1.96\frac{\sqrt{37.2}}{5}$$

Therefore, the 95% interval is

$$\left[9.6 - 1.96\frac{\sqrt{37.2}}{5}, 9.6 + 1.96\frac{\sqrt{37.2}}{5}\right]$$

Set 33 Suppose that you have tossed a coin – which may or may not be balanced – five times and reported the outcomes 1, 0, 0, 1, and 1, with Heads coded 1's and Tails by 0's.

(a) Calculate the sample proportion $\hat{p}$ of Heads.

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{x_i = H}$$

$$\hat{p} = \frac{3}{5}$$

(b) Construct 95% (asymptotic) interval for the population probability of Heads.

The equation for a confidence interval with 95% confidence is,

$$\hat{p} - 1.96\frac{\hat{\sigma}}{\sqrt{n}} \leq p \leq \hat{p} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}$$

And the standard deviation of the population sample is,

$$\hat{\sigma} = \hat{p}(1 - \hat{p})$$

Then,

$$\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \leq p \leq \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\frac{3}{5} - 1.96\frac{\sqrt{\frac{3}{5}\left(1 - \frac{3}{5}\right)}}{\sqrt{5}} \leq p \leq \frac{3}{5} + 1.96\frac{\sqrt{\frac{3}{5}\left(1 - \frac{3}{5}\right)}}{\sqrt{5}}$$

$$\frac{3}{5} - 1.96\frac{\sqrt{\frac{6}{25}}}{\sqrt{5}} \leq p \leq \frac{3}{5} + 1.96\frac{\sqrt{\frac{6}{25}}}{\sqrt{5}}$$

$$\frac{3}{5} - 1.96\frac{\sqrt{6}}{5\sqrt{5}} \leq p \leq \frac{3}{5} + 1.96\frac{\sqrt{6}}{5\sqrt{5}}$$

Therefore, the 95% confidence interval is

$$\left[\frac{3\sqrt{5} - 1.96\sqrt{6}}{5\sqrt{5}}, \frac{3\sqrt{5} + 1.96\sqrt{6}}{5\sqrt{5}}\right]$$

(c) Construct a conservative 95% (asymptotic) confidence interval for the population probability of Heads.

The conservative assumption is

$$\hat{p} = 0.5$$

Then,

$$\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \leq p \leq \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\frac{1}{2} - 1.96\frac{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}}{\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}}{\sqrt{5}}$$

$$\frac{1}{2} - 1.96\frac{\sqrt{\frac{1}{4}}}{\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{\sqrt{\frac{1}{4}}}{\sqrt{5}}$$

$$\frac{1}{2} - 1.96\frac{1}{2\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{1}{2\sqrt{5}}$$

Therefore, the 95% conservative confidence interval is

$$\left[\frac{\sqrt{5} - 1.96}{2\sqrt{5}}, \frac{\sqrt{5} + 1.96}{2\sqrt{5}}\right]$$

Set 34 Suppose that you own a wind farm with five wind turbines of the same type, and suppose that you have kept the records of their first failure times ${t_1}^{obs}, \dots, {t_5}^{obs}$, which are 14, 6, 8, 11, and 9 months.

(a) Construct a 95% (asymptotic) confidence interval for the probability that the wind turbine of this type functions without a failure for $t = 10$ months after its installation.

$$\hat{p} = \mathbb{P}(T \geq 10) = \frac{2}{5}$$

Then,

$$\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \leq p \leq \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\frac{2}{5} - 1.96\frac{\sqrt{\frac{2}{5}\left(1-\frac{2}{5}\right)}}{\sqrt{5}} \leq p \leq \frac{2}{5} + 1.96\frac{\sqrt{\frac{2}{5}\left(1-\frac{2}{5}\right)}}{\sqrt{5}}$$

$$\frac{2}{5} - 1.96\frac{\sqrt{\frac{6}{25}}}{\sqrt{5}} \leq p \leq \frac{2}{5} + 1.96\frac{\sqrt{\frac{6}{25}}}{\sqrt{5}}$$

$$\frac{2}{5} - 1.96\frac{\sqrt{6}}{5\sqrt{5}} \leq p \leq \frac{2}{5} + 1.96\frac{\sqrt{6}}{5\sqrt{5}}$$

Therefore, the 95% confidence interval is

$$\left[\frac{2\sqrt{5} - 1.96\sqrt{6}}{5\sqrt{5}}, \frac{2\sqrt{5} + 1.96\sqrt{6}}{5\sqrt{5}}\right]$$

(b) Construct a *conservative* 95% (asymptotic) confidence interval for the probability that the wind turbine of this type functions without a failure for $t = 10$ months after its installation.

The conservative assumption is

$$\hat{p} = 0.5$$

Then,

$$\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \leq p \leq \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\frac{1}{2} - 1.96\frac{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}}{\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}}{\sqrt{5}}$$

$$\frac{1}{2} - 1.96\frac{\sqrt{\frac{1}{4}}}{\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{\sqrt{\frac{1}{4}}}{\sqrt{5}}$$

$$\frac{1}{2} - 1.96\frac{1}{2\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{1}{2\sqrt{5}}$$

Therefore, the 95% conservative confidence interval is

$$\left[\frac{\sqrt{5} - 1.96}{2\sqrt{5}}, \frac{\sqrt{5} + 1.96}{2\sqrt{5}}\right]$$

(c) Construct a 95% (asymptotic) confidence interval for the probability that the wind turbine of this type fails within $t = 10$ months after its installation.

$$\hat{p} = \mathbb{P}(T < 10) = 1 - \mathbb{P}(T \geq 10) = 1 - \frac{2}{5} = \frac{3}{5}$$

Then,

$$\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \leq p \leq \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\frac{3}{5} - 1.96\frac{\sqrt{\frac{3}{5}\left(1-\frac{3}{5}\right)}}{\sqrt{5}} \leq p \leq \frac{3}{5} + 1.96\frac{\sqrt{\frac{3}{5}\left(1-\frac{3}{5}\right)}}{\sqrt{5}}$$

$$\frac{3}{5} - 1.96\frac{\sqrt{\frac{6}{25}}}{\sqrt{5}} \leq p \leq \frac{3}{5} + 1.96\frac{\sqrt{\frac{6}{25}}}{\sqrt{5}}$$

$$\frac{3}{5} - 1.96\frac{\sqrt{6}}{5\sqrt{5}} \leq p \leq \frac{3}{5} + 1.96\frac{\sqrt{6}}{5\sqrt{5}}$$

Therefore, the 95% confidence interval is

$$\left[\frac{3\sqrt{5} - 1.96\sqrt{6}}{5\sqrt{5}}, \frac{3\sqrt{5} + 1.96\sqrt{6}}{5\sqrt{5}}\right]$$

(d) Construct a *conservative* 95% (asymptotic) confidence interval for the probability that the wind turbine of this type fails within for $t = 10$ months after its installation.

The conservative assumption is

$$\hat{p} = 0.5$$

Then,

$$\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \leq p \leq \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\frac{1}{2} - 1.96\frac{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}}{\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}}{\sqrt{5}}$$

$$\frac{1}{2} - 1.96\frac{\sqrt{\frac{1}{4}}}{\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{\sqrt{\frac{1}{4}}}{\sqrt{5}}$$

$$\frac{1}{2} - 1.96\frac{1}{2\sqrt{5}} \leq p \leq \frac{1}{2} + 1.96\frac{1}{2\sqrt{5}}$$

Therefore, the 95% conservative confidence interval is

$$\left[\frac{\sqrt{5}-1.96}{2\sqrt{5}}, \frac{\sqrt{5}+1.96}{2\sqrt{5}}\right]$$

Set 35 Suppose that you have designed a product and want to assess its reliability in the form of its functionality during the required, say, 10 day mission. For this, you want to have an estimate of the reliability function $R(t)$ at $t = 10$, and you also want to have a 95% (asymptotic) confidence interval for $R(10)$ whose margin of error does not exceed 0.01. How many prototypes should you manufacture and then wait for their failure times in order to construct the aforementioned confidence interval when:

(a) there is no additional information available to you?

Since there is no $\hat{p}$ provided, choose a conservative $\hat{p} = 0.5$

Next, since the margin of error is,

$$Margin\ of\ error = z * SE$$

And,

$$SE(standard\ error) = \frac{\sigma}{\sqrt{n}}$$

$$z = 1.96$$

Then,

$$\frac{1}{100} = 1.96\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = 100 * 1.96 * \sqrt{\hat{p}(1-\hat{p})}$$

$$\sqrt{n} = 100 * 1.96 * \sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}$$

$$\sqrt{n} = 100 * 1.96 * \frac{1}{2}$$

$$\sqrt{n} = 25 * 1.96$$

$$n = 49^2$$

Therefore, $n = 2401$

(b) your expertise in physics, chemistry, and of course engineering tells you that $R(10)$ should be somewhere between 0.85 and 0.98?

Since $0.85 \leq R(10) \leq 0.98$ then select a $\hat{p}$ value that is closest to 0.5, $\hat{p} = 0.85$,

$$\frac{1}{100} = 1.96\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = 100 * 1.96 * \sqrt{\hat{p}(1-\hat{p})}$$

$$\sqrt{n} = 100 * 1.96 * \sqrt{\frac{85}{100}\left(1 - \frac{85}{100}\right)}$$

$$\sqrt{n} = 100 * 1.96 * \frac{85}{100} * \frac{15}{100}$$

$$\sqrt{n} = 12.75 * 1.96$$

$$n = 24.99^2$$

Therefore, $n \approx 624.50$

(c) your expertise tells you that $R(10)$ should be somewhere between 0.45 and 0.70?

Since $0.45 \leq R(10) \leq 0.70$ then select a $\hat{p}$ value that is closest to 0.5, $\hat{p} = 0.5$,

$$\frac{1}{100} = 1.96\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = 100 * 1.96 * \sqrt{\hat{p}(1-\hat{p})}$$

$$\sqrt{n} = 100 * 1.96 * \sqrt{\frac{1}{2}\left(1 - \frac{1}{2}\right)}$$

$$\sqrt{n} = 100 * 1.96 * \frac{1}{2}$$

$$\sqrt{n} = 25 * 1.96$$

$$n = 49^2$$

Therefore, $n = 2401$

Set 36 Suppose that you own a fleet of taxis consisting of hybrid cars of the same type, and suppose that you have decided to re-invest some of your profits by adding solar roofs to the taxis. Hence, you buy the best solar cells available on the market and, being an innovative engineer, decide to arrange them in a most efficient way. You quickly realize that the shape of the car roof is one of the most important factors, and you can easily shape the roof into any form of your liking (the drag coefficient is not an issue for your taxis because they cannot drive fast in the city). Your computer modeling suggests a certain shape, which is somewhat different from the original roof-shape of the taxis. Should you keep the original roof-shape or modify it? To answer this question, you randomly select ten taxis from your fleet, add solar roofs to all of them, drive for a week, and record their gas-free mileage. After that, you modify the roofs of the same ten taxis, add solar roofs, drive them for a week, and record their gas-free mileage. Hence, you have ten pairs of data. Construct a 95% (asymptotic) confidence interval for the difference between the (population) average gas-free mileage under the original and modified roofs.

Let X be an RV that determines the average gas free mileage under the original roof. Let Y be an RV that determines the average gas free mileage under the modified roof. Let Z be an RV that determines the difference between gas free mileage under original vs. modified roofs, where $z_i = x_i - y_i$

$$\bar{Z} = \mathbb{E}[Z]$$

$$\hat{\sigma}_Z = \sqrt{\mathbb{E}[(\bar{Z} - Z)^2]}$$

Then,

$$\bar{Z} - 1.96\frac{\hat{\sigma}_Z}{\sqrt{n}} \le \mu \le \bar{Z} + 1.96\frac{\hat{\sigma}_Z}{\sqrt{n}}$$

$$\bar{Z} - 1.96\frac{\hat{\sigma}_Z}{\sqrt{10}} \le \mu \le \bar{Z} + 1.96\frac{\hat{\sigma}_Z}{\sqrt{10}}$$

Therefore, the mean difference confidence interval is

$$\left[\bar{Z} - 1.96\frac{\hat{\sigma}_Z}{\sqrt{10}}, \bar{Z} + 1.96\frac{\hat{\sigma}_Z}{\sqrt{10}}\right]$$

Set 37 Suppose that instead of using ten same cars as in Set 36, you are now dealing with a taxi fleet that is sufficiently large to allow you to select a number of cars, say 10, whose unmodified roofs are fitted with solar panels, and also select a number of cars, say 15, whose modified roofs are fitted with solar panels. Hence, the whole experiment can now be run within one week, instead of two as in Set 36, and so you obtain two sets of data at the same time: one has 10 and another 15 observations (gas-free mileage recordings). Construct a 95% (asymptotic) confidence interval for the difference between the (population) average gas-free mileage under the original and modified roofs.

Let X be an RV that determines the average gas free mileage under the original roof.
Let Y be an RV that determines the average gas free mileage under the modified roof.

$$\bar{x} = \mathbb{E}[X], \bar{y} = \mathbb{E}[Y]$$

$$\hat{\sigma}_x = \sqrt{\mathbb{E}[(\bar{x} - x)^2]}, \hat{\sigma}_y = \sqrt{\mathbb{E}[(\bar{y} - y)^2]}$$

Then,

$$\bar{x} - \bar{y} - 1.96\left(\frac{\hat{\sigma}_x}{\sqrt{n_x}} + \frac{\hat{\sigma}_y}{\sqrt{n_y}}\right) \leq \mu \leq \bar{x} - \bar{y} + 1.96\left(\frac{\hat{\sigma}_x}{\sqrt{n_x}} + \frac{\hat{\sigma}_y}{\sqrt{n_y}}\right)$$

$$\bar{x} - \bar{y} - 1.96\left(\frac{\hat{\sigma}_x}{\sqrt{10}} + \frac{\hat{\sigma}_y}{\sqrt{15}}\right) \mu \leq \bar{x} - \bar{y} + 1.96\left(\frac{\hat{\sigma}_x}{\sqrt{10}} + \frac{\hat{\sigma}_y}{\sqrt{15}}\right)$$

Therefore, the mean difference confidence interval is

$$\left[\bar{x} - \bar{y} - 1.96\left(\frac{\hat{\sigma}_x}{\sqrt{10}} + \frac{\hat{\sigma}_y}{\sqrt{15}}\right), \bar{x} - \bar{y} + 1.96\left(\frac{\hat{\sigma}_x}{\sqrt{10}} + \frac{\hat{\sigma}_y}{\sqrt{15}}\right)\right]$$

Set 38 Let $x_1{}^{obs}, \ldots, x_n{}^{obs}$ be observations, and let x and $\hat{\sigma}^2$ be the sample mean and variance, respectively. Does the 95% confidence interval

$$\left[\bar{x} - 1.96\frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

(a) cover the (unknown) population mean μ?

The confidence interval covers the population mean with a confidence percentage of 95%.

(b) cover the sample mean $\bar{x}$?

The confidence interval covers the sample mean.

Set 39 Suppose you have paired data $(x_1{}^{obs}, y_1{}^{obs}), \dots, (x_n{}^{obs}, y_n{}^{obs})$ which are outcomes of two random variables X and Y. You construct a 95% confidence interval for the difference of the means of X and Y.

(a) Assuming that the correlation between X and Y is positive, is your constructed confidence interval shorter or longer than the 95% confidence interval that you would have in the case of independent X and Y? Prove your answer.

A positive correlation will mean that the confidence interval range will be shorter when compared to a confidence interval of independent variables because the positive correlation means that both random variables correlate in the same direction which means that the difference of those random variables will be lower, hence, tighter confidence interval.

(b) If the 95% confidence interval does not cover 0, would you retain or reject someone's claim that the means of X and Y are equal? Justify your answer.

If the confidence interval covers 0 then the equivalence claim for the means of X and Y cannot be rejected because the possibility of a difference mean of 0, no difference, will still hold.

(c) If the entire 95% confidence interval is to the right of 0, would you conclude or not that the mean of X is larger than the mean of Y? Justify your answer.

With a 95% confidence, yes. This is with the assumption that the means difference confidence interval is computed as $X - Y$.

Set 40 Least-squares regression line

(a) Derive expressions for a and b that give the minimal value of the two-argument function defined by $g(a,b) = \mathbb{E}[(a + bX - Y)^2]$. (Hint: a is called the intercept and b the slope of the least-squares regression line, which is $y = a + bx$.)

$$g(a,b) = \mathbb{E}[(a + bX - Y)^2]$$

$$g(a,b) = \sum_{i=1}^{n} (a + bx_i - y_i)^2$$

Take the partial derivative with respect to a and find its min value

$$\sum_{i=1}^{n} (a + bx_i - y_i) = 0$$

$$\sum_{i=1}^{n} a + \sum_{i=1}^{n} bx_i = \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$$

$$\frac{1}{n}\sum_{i=1}^{n} y_i = a + \frac{b}{n}\sum_{i=1}^{n} x_i$$

$$\mathbb{E}[Y] = a + \mathbb{E}[X]$$

$$a = \mathbb{E}[Y] - b\mathbb{E}[X]$$

$$a = \bar{Y} - b\bar{X}$$

Take the partial derivative with respect to b and find its min value

$$g(a,b) = \sum_{i=1}^{n} (\bar{Y} - b\bar{X} + bx_i - y_i)^2$$

$$g(a,b) = \sum_{i=1}^{n} \left((\bar{Y} - y_i) + b(x_i - \bar{X})\right)^2$$

$$\sum_{i=1}^{n}[(\bar{Y}-y_i)+b(x_i-\bar{X})](x_i-\bar{X})=0$$

$$\sum_{i=1}^{n}(\bar{Y}-y_i)(x_i-\bar{X})+b(x_i-\bar{X})^2=0$$

$$-\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)+b\sum_{i=1}^{n}(\bar{X}-x_i)^2=0$$

$$b\sum_{i=1}^{n}(\bar{X}-x_i)^2=\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)$$

$$b=\frac{\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)}{\sum_{i=1}^{n}(\bar{X}-x_i)^2}$$

Finally, substitute b in to find a

$$a=\bar{Y}-b\bar{X}$$

$$a=\bar{Y}-\frac{\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)}{\sum_{i=1}^{n}(\bar{X}-x_i)^2}\bar{X}$$

(b) Construct empirical estimators $\hat{a}$ and $\hat{b}$ for a and b, respectively, and then compute the estimators and draw the corresponding least-squares regression line using the following data:

Year(i)	1	2	3	4	5	6	7	8	9	10	11	12
Market(x_i)	0.15	0.13	0.07	0.12	-0.04	0.31	0.23	0.31	0.02	-0.07	0.07	0.02
Fund(y_i)	-0.05	0.05	0.01	0.25	0.04	0.15	0.40	0.29	0.33	-0.03	0.02	-0.02

$$\bar{X}=\frac{15+13+7+12-4+31+23+31+2-7+7+2}{100*12}=0.11$$

$$\bar{Y}=\frac{-5+5+1+25+4+15+40+29+33-3+2-2}{100*12}=0.12$$

$$\begin{aligned}\mathrm{cov}(X,Y)=&[(12+5)*(11-15)+(12-5)*(11-13)+(12-1)*(11-7)\\&+(12-25)*(11-12)+(12-4)*(11+4)+(12-15)*(11-31)\\&+(12-40)*(11-23)+(12-29)*(11-31)+(12-33)*(11-2)\\&+(12+3)*(11+7)+(12-2)*(11-7)+(12+2)*(11-2)]\div 100\\&=0.1078\end{aligned}$$

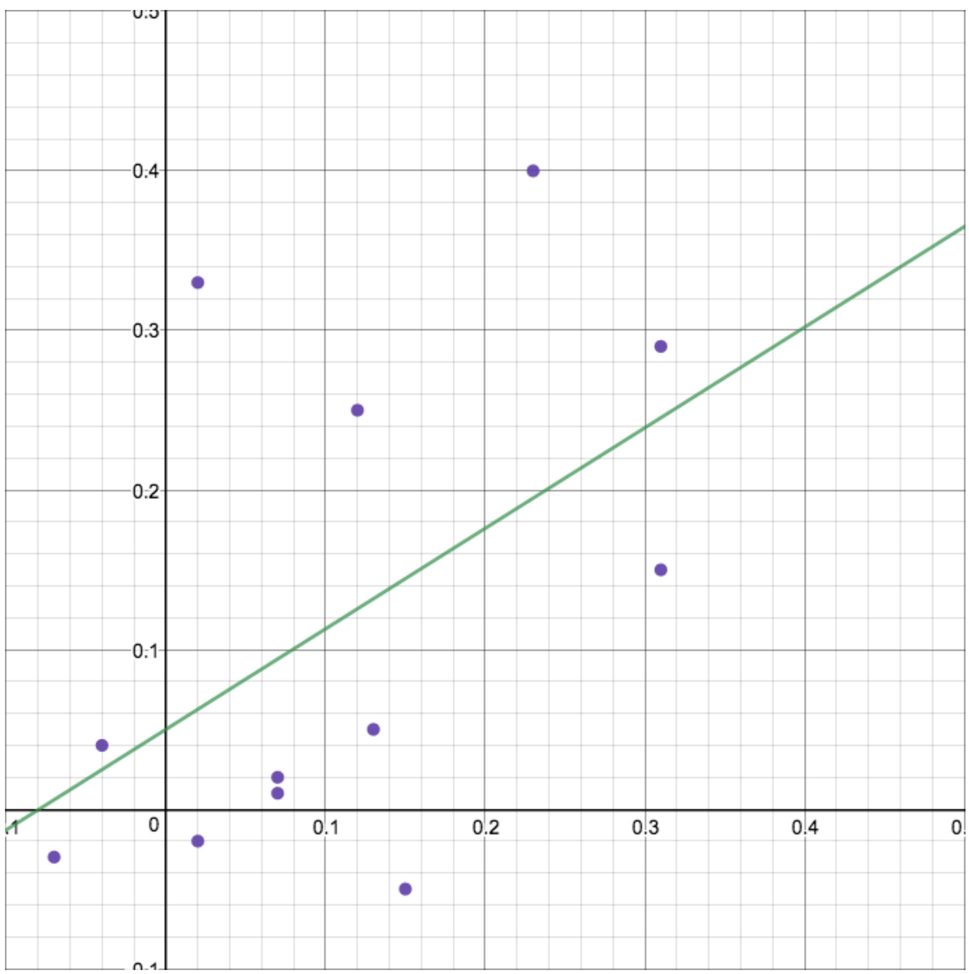
$$\begin{aligned}\operatorname{var}(X) = [&(11-15)^2 + (11-13)^2 + (11-7)^2 + (11-12)^2 + (11+4)^2 + (11-31)^2 \\ &+ (11-23)^2 + (11-31)^2 + (11-2)^2 + (11+7)^2 + (11-7)^2 \\ &+ (11-2)^2] \div 100^2 = 0.1708\end{aligned}$$

Therefore,

$$b = \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X,Y)} = \frac{0.1078}{0.1708} \approx 0.63$$

$$a = \bar{Y} - \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X,Y)}\bar{X} = 0.12 - \frac{0.1078}{0.1708}0.11 \approx 0.05$$

$$y \approx 0.05 + 0.63x$$



(c) Prove the equation

$$\frac{\text{cov}(X,Y)}{\sigma_X^2} = \text{corr}(X,Y)\frac{\sigma_Y}{\sigma_X}$$

where $\text{cov}(X,Y)$ is the covariance between X and Y, and $\text{corr}(X,Y)$ is the correlation between X and Y. (Note: The left-hand side of equation is the famous "beta" that every financial portfolio manager knows and uses on the daily basis, with X being the market return and Y the fund return.)

$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Therefore,

$$\frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}\frac{\sigma_Y}{\sigma_X}$$

$$\frac{\text{cov}(X,Y)}{\sigma_X}\frac{1}{\sigma_X}$$

$$\frac{\text{cov}(X,Y)}{\sigma_X^2}$$

(d) Prove that the mean squared error $MSE := \mathbb{E}\left[\left(\hat{Y} - Y\right)^2\right]$ between the least squares predictor $\hat{Y} = a + bX$ of the response variable Y is equal to $\sigma_Y^2(1-\rho^2)$ where σ_Y^2 is the variance of Y and $\rho = \text{corr}(X,Y)$ is the correlation between X and Y.

$$\mathbb{E}\left[\left(\hat{Y} - Y\right)^2\right]$$

$$\hat{Y} = a + bX$$

$$\hat{Y} = \bar{Y} - \frac{\text{cov}(X,Y)}{\sigma_X}\bar{X} + \frac{\text{cov}(X,Y)}{\sigma_X}X$$

$$\mathbb{E}\left[\left(\mathbb{E}[Y] - \mathbb{E}[X]\frac{\text{cov}(X,Y)}{\sigma_X} + X\frac{\text{cov}(X,Y)}{\sigma_X} - Y\right)^2\right]$$

$$\mathbb{E}\left[\mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + 2X\mathbb{E}[Y]\frac{\operatorname{cov}(X,Y)}{\sigma_X} - 2Y\mathbb{E}[Y] + \mathbb{E}[X]^2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} - 2X\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} + 2Y\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + X^2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} - 2XY\frac{\operatorname{cov}(X,Y)}{\sigma_X} + Y^2\right]$$

$$\mathbb{E}[\mathbb{E}[Y]^2] + \mathbb{E}\left[-2\mathbb{E}[X]\mathbb{E}[Y]\frac{\operatorname{cov}(X,Y)}{\sigma_X}\right] + \mathbb{E}\left[2X\mathbb{E}[Y]\frac{\operatorname{cov}(X,Y)}{\sigma_X}\right] + \mathbb{E}\left[-2Y\mathbb{E}[Y]\right] + \mathbb{E}\left[\mathbb{E}[X]^2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2}\right] + \mathbb{E}\left[-2X\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2}\right] + \mathbb{E}\left[2Y\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)}{\sigma_X}\right] + \mathbb{E}\left[X^2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2}\right] + \mathbb{E}\left[-2XY\frac{\operatorname{cov}(X,Y)}{\sigma_X}\right] + \mathbb{E}[Y^2]$$

$$\mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + 2\mathbb{E}[X]\mathbb{E}[Y]\frac{\operatorname{cov}(X,Y)}{\sigma_X} - 2\mathbb{E}[Y]\mathbb{E}[Y] + \mathbb{E}[X]^2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} - 2\mathbb{E}[X]\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} + 2\mathbb{E}[Y]\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + \mathbb{E}[X^2]\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} - 2\mathbb{E}[XY]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + \mathbb{E}[Y^2]$$

$$\mathbb{E}[Y]^2 - 2\mathbb{E}[Y]^2 - \mathbb{E}[X]^2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} + 2\mathbb{E}[Y]\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + \mathbb{E}[X^2]\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} - 2\mathbb{E}[XY]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + \mathbb{E}[Y^2]$$

$$\mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + \mathbb{E}[X^2]\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} - \mathbb{E}[X]^2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2} - 2\mathbb{E}[XY]\frac{\operatorname{cov}(X,Y)}{\sigma_X} + 2\mathbb{E}[Y]\mathbb{E}[X]\frac{\operatorname{cov}(X,Y)}{\sigma_X}$$

$$\mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + (\mathbb{E}[X^2] - \mathbb{E}[X]^2)\left(\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2}\right) - 2\left(\frac{\operatorname{cov}(X,Y)}{\sigma_X}\right)(\mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X])$$

$$\sigma_Y^2 + \sigma_X^2\left(\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2}\right) - 2\left(\frac{\operatorname{cov}(X,Y)}{\sigma_X}\right)\operatorname{cov}(X,Y)$$

$$\sigma_Y^2 + \operatorname{cov}(X,Y)^2 - 2\frac{\operatorname{cov}(X,Y)^2}{\sigma_X}$$

$$\sigma_Y^2 + \mathrm{cov}(X,Y)^2\left(1-\frac{2}{\sigma_X}\right)$$

$$\sigma_Y^2 + \mathrm{cov}(X,Y)^2\left(\frac{\sigma_X}{\sigma_X}-\frac{2}{\sigma_X}\right)$$

$$\sigma_Y^2 + \mathrm{cov}(X,Y)^2\left(\frac{\sigma_X-2}{\sigma_X}\right)$$

$$\sigma_Y^2 + \mathrm{cov}(X,Y)^2\left(\frac{\sigma_X-2}{\sigma_X}\right)\left(\frac{\sigma_X}{\sigma_X}\right)$$

$$\sigma_Y^2 + \mathrm{cov}(X,Y)^2\sigma_X\left(\frac{\sigma_X-2}{\sigma_X^2}\right)\left(\frac{\sigma_Y^2}{\sigma_Y^2}\right)$$

$$\sigma_Y^2\left(1 + \mathrm{cov}(X,Y)^2\sigma_X\left(\frac{\sigma_X-2}{\sigma_X^2\sigma_Y^2}\right)\right)$$

$$\sigma_Y^2\left(1 + \sigma_X(\sigma_X-2)\left(\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2\sigma_Y^2}\right)\right)$$

$$\sigma_Y^2(1+(\sigma_X^2-2\sigma_X)\rho^2)$$

Doesn’t work…

Set 41 Let X have the uniform on $[-1, 1]$ density, and let Y be another random variable given by the equation $Y = X^2$. Hence, the value of Y is completely determined by the value of X. Are the two random variables X and Y correlated or uncorrelated? Prove your answer.

$$\operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$\operatorname{cov}(X, Y) = \operatorname{cov}(X, X^2)$$

$$\mathbb{E}[(X - \mathbb{E}[X])(X^2 - \mathbb{E}[X^2])]$$

$$\mathbb{E}\left[X^3 - X\mathbb{E}[X^2] - X^2\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[X^2]\right]$$

$$\mathbb{E}[X^3] + \mathbb{E}\left[-X\mathbb{E}[X^2]\right] + \mathbb{E}\left[-X^2\mathbb{E}[X]\right] + \mathbb{E}\left[\mathbb{E}[X]\mathbb{E}[X^2]\right]$$

$$\mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] - \mathbb{E}[X]\mathbb{E}[X^2] + \mathbb{E}[X]\mathbb{E}[X^2]$$

$$\mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2]$$

$$\int_{-1}^{1} x^3 dx - \int_{-1}^{1} x^2 dx * \int_{-1}^{1} x^3 dx = 0$$

Therefore, since the covariance is 0, then the correlation is also 0. Hence, uncorrelated.

Set 42 Suppose that you look after a wind farm with 20 wind turbines, among which 5 turbines are in location A and the remaining 15 turbines are in another location B. Calculate the mean time to failure of the entire wind farm when you only know that the mean time to failure of the turbines in location A is 6 months, and in location B is 4 months.

$$MTTF = \mathbb{E}[T|A]\mathbb{P}(A) + \mathbb{E}[T|B]\mathbb{P}(B)$$

$$MTTF = 6 * \frac{5}{20} + 4 * \frac{15}{20}$$

$$MTTF = \frac{3}{2} + 3$$

$$MTTF = 4.5$$

Set 43 Suppose that you own a grocery store. Let N denote the number of shoppers who visit your grocery store, and let X_n (where $n = 1, 2 \ldots$) denote the amount of money that the n-th shopper spends during one week. Based on historical data, you know that each shopper spends, in average, \$200 in your store during one week, and the average number of shoppers who visit your store during one week is 500. Let S_N be the total sum of money spent by all the N shoppers during one week. Assume that the number of shoppers N and the amounts $X_n, n \geq 1$, of money that they spend are independent. What is the average $\mathbb{E}[S_N]$ of the total amount of money S_N spent during one week at your grocery store?

$$S_N = \sum_{i=1}^{n} X_i$$

$$\mathbb{E}[S_N] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]$$

$$\mathbb{E}[S_N] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

Since,

$$\mathbb{E}[X_i] = \mathbb{E}[S_i | N = i]\mathbb{P}(N = i)$$

Then,

$$\mathbb{E}[S_N] = \sum_{i=1}^{n} \mathbb{E}[S_i | N = i] * \mathbb{P}(N = i)$$

We know that the average customer spends \$200,

$$\mathbb{E}[S_N] = \sum_{i=1}^{n} 200 * \mathbb{P}(N = i)$$

$$\mathbb{E}[S_N] = 200 \sum_{i=1}^{n} \mathbb{P}(N = i)$$

$$\mathbb{E}[S_N] = 200 * 500$$

$$\mathbb{E}[S_N] = 100{,}000$$

Set 44 Let X be a discrete random variable, and let A and B be two disjoint (that is, $A \cap B = \emptyset$) and exhaustive (that is, $A \cup B = \Omega$) subsets of the sample space Ω. Prove

$$\mathbb{E}[X] = \mathbb{E}[X|A]\mathbb{P}(A) + \mathbb{E}[X|B]\mathbb{P}(B)$$

The law of total expectation states that,

$$\mathbb{E}[X] = \sum_{i}^{n} \mathbb{E}[X|Y_i] * \mathbb{P}(Y_i)$$

Where $Y_1, Y_2, \dots, Y_n$ are portions of the whole space Ω. Next, we know that A and B are the partitions that form Ω, therefore,

$$\mathbb{E}[X] = \sum_{i}^{n} \mathbb{E}[X|Y_i] * \mathbb{P}(Y_i)$$

$$\mathbb{E}[X] = \mathbb{E}[X|Y_0] * \mathbb{P}(Y_0) + \mathbb{E}[X|Y_1] * \mathbb{P}(Y_1)$$

$$\mathbb{E}[X] = \mathbb{E}[X|A] * \mathbb{P}(A) + \mathbb{E}[X|B] * \mathbb{P}(B)$$

Set 45 Let X and Y be two discrete random variables. Prove that the expectation of the conditional expectation is the unconditional expectation, that is,

$$\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \mathbb{E}[X]$$

$$\mathbb{E}\left[\sum_x x * \mathbb{P}(X = x|Y)\right]$$

$$\sum_y \left[\sum_x x * \mathbb{P}(X = x|Y = y)\right] * \mathbb{P}(Y = y)$$

$$\sum_y \sum_x x * \mathbb{P}(X = x|Y = y) * \mathbb{P}(Y = y)$$

$$\sum_x x \sum_y \mathbb{P}(X = x|Y = y) * \mathbb{P}(Y = y)$$

$$\sum_x x \sum_y \mathbb{P}(X = x|Y = y)$$

$$\sum_x x * \mathbb{P}(X = x)$$

$$\mathbb{E}[X]$$