Math 218, Mathematical Statistics
D Joyce, Spring 2016

Chap. 6, p. 229, exercises 1, 2ab, 5, 7, 11, 13a, 14, 15ab.

Selected answers.

1. State whether each boldfaced number is a parameter or a statistic.

These take a little bit of interpretation. You need to see what the stated sample is a sample of. If the number depends only on the total population, it's a parameter, but if it depends on the sample, it's a statistic.

a. A shipment of 1000 fuses contains **3** defective fuses. A sample of 25 fuses contained **0** defectives.

The sample is from a population of size 1000. The number **0** of defective fuses in a sample of size $n = 25$ is a statistic. The number **3** of defective fuses in the whole population is a parameter of the distribution.

b. The speed of 100 vehicles was monitored. It was found that **63** vehicles exceeded the posted speed limit.

The sample has size $n = 100$ of some unknown sized population. The number **63** is a statistic of that sample.

c. A telephone poll of registered voters one week before a statewide election showed that **48%** would vote for the current governor, who was running for reelection. The final election returns showed that the incumbent won with **52%** of the votes cast.

The **48%** depends on the sample, so it's a statistic of the sample. The **52%** depends on the entire voting population, so it's a paramater.

5. Suppose we have n independent Bernoulli trials with true success probability p. Consider two estimators of p: $\hat{p}_1 = \hat{p}$ where $\hat{p}$ is the sample proportion of successes and $\hat{p}_2 = \frac{1}{2}$, a fixed constant.

a. Find the expected value and bias of each estimator.
$E(\hat{p}_1) = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}(p + \cdots + p) = p$.
$E(\hat{p}_2) = E(\frac{1}{2}) = \frac{1}{2}$.
$\text{Bias}(\hat{p}_1) = E(\hat{p}_1) - p = p - p = 0$. It's unbiased.
$\text{Bias}(\hat{p}_2) = E(\hat{p}_2) - p = \frac{1}{2} - p$. It's biased unless $p = \frac{1}{2}$.

b. Find the variance of each estimator. Which estimator has the lower variance?

$$\begin{aligned}\text{Var}(\hat{p}_1) &= \text{Var}(\frac{1}{n}(X_1 + \cdots + X_n)) \\ &= \frac{1}{n^2}\text{Var}(X_1 + \cdots + X_n) \\ &= \frac{1}{n^2}(\text{Var}(X_1) + \cdots + \text{Var}(X_n)) \\ &= \frac{1}{n^2}(pq + \cdots + pq) \\ &= \frac{1}{n}pq\end{aligned}$$

$\text{Var}(\hat{p}_2) = \text{Var}(\frac{1}{2}) = 0$.
Therefore, $\hat{p}_2$ has the smaller variance.

c. Find the MSE (mean squared error) of each estimator and compare them by plotting against the true p. Use $n = 4$. Comment on the comparison.

$\text{MSE}(\hat{p}_1) = \text{Var}(\hat{p}_1) + (\text{Bias}(\hat{p}_1))^2 = \frac{1}{n}pq + 0 = \frac{1}{n}pq = \frac{1}{n}p(1-p)$.
$\text{MSE}(\hat{p}_2) = \text{Var}(\hat{p}_2) + (\text{Bias}(\hat{p}_2))^2 = 0 + (\frac{1}{2} - p)^2 = (\frac{1}{2} - p)^2$.
When $n =$, $\text{MSE}(\hat{p}_1) = \frac{1}{4}p(1-p)$, while $\text{MSE}(\hat{p}_2) = (\frac{1}{2} - p)^2$.

The graph of $\text{MSE}(\hat{p}_1)$ is a parabola passing through the points $(0,0)$, $(\frac{1}{2}, \frac{1}{16})$, and $(1,0)$.

The graph of $\text{MSE}(\hat{p}_2)$ is also a parabola, but it passes through the points $(0, \frac{1}{4})$, $(\frac{1}{2}, 0)$, and $(1, \frac{1}{4})$.

Thus, for most values of p, $\hat{p}_1$ has a smaller MSE, but if p happens to be near $\frac{1}{2}$, then $\hat{p}_2$ has a smaller MSE.

11. Consider the probability

$$P\left(-1.645 \leq Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq +1.645\right)$$

where $\overline{X}$ is the sample mean of a random sample of size n drawn from a normal distribution with mean μ and variance σ^2.

a. Use this statement to find a confidence interval for μ. What is the confidence level of this confidence interval?

The statistic $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ is a standard normal distribution assuming X has a normal distribution as stated. From the table for a standard normal distribution, $\Phi(1.645) = 0.95$, so $P(-1.645 \leq Z \leq +1.645) = 0.90$. Therefore, the confidence level here is 90%. The interval $[-1.645, 1, 645]$ is for $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$. So, for μ, the interval is

$$\left[\overline{X} - 1.645\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.645\frac{\sigma}{\sqrt{n}}\right].$$

b. A sample of $n = 100$ is taken from a normal population with $\sigma = 10$. The sample mean is 30. Calculate the confidence interval for μ using the result from part a.

Since $n = 100$, $\overline{X} = 30$ and $\sigma = 10$, the above interval is

$$\left[30 - 1.645\frac{10}{\sqrt{100}}, 30 + 1.645\frac{10}{\sqrt{100}}\right],$$

which simplifies to $[28.355, 31.645]$.

c. What is the probability that μ is included in the confidence interval calculated in part b?

Although $P(\overline{X} - 1.645 \leq \mu \leq \overline{X} + 1.645) = 0.90$, the probability $P(28.355 \leq \mu \leq 31.645)$ is either equal to 0 or to 1 depending on whether the constant μ lies in the interval or not.

13a. Suppose that 100 random samples of size 9 are generated from a normal distribution with mean $\mu = 70$ and variance $\sigma^2 = 3^2$ and the associated 95% confidence interval is calculated for each sample.

How many of these 100 intervals would you expect to contain the true $\mu = 70$?

That would be 95% of the 100 intervals, that is, 95 intervals, to contain μ.

14. A random sample of size 25 from a normal distribution with mean μ and variance $\sigma^2 = 6^2$ has a mean of $\overline{x} = 16.3$.

a. Calculate confidence intervals for μ for three levels of confidence: 80%, 90%, and 99%. How do the confidence widths change?

The widths of the interval will increase as the confidence level increases. The interval for confidence level $1 - \alpha$ is

$$\left[\overline{X} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right].$$

For an 80% confidence level, $\alpha/2 = 0.1$, so, by the tables, $z_{\alpha/2} = 1.282$, so $z_{\alpha/2}\sigma/\sqrt{n} = 1.538$, and, therefore,the confidence interval is

$$[16.3 - 1.538, 16.3 + 1.538] = [14.76, 17.84].$$

For an 90% confidence level, $\alpha/2 = 0.05$, so $z_{\alpha/2} = 1.645$, $z_{\alpha/2}\sigma/\sqrt{n} = 1.974$, and the confidence interval is

$$[16.3 - 1.974, 16.3 + 1.974] = [14.33, 18.27].$$

For an 99% confidence level, $\alpha/2 = 0.005$, so $z_{\alpha/2} = 2.575$, $z_{\alpha/2}\sigma/\sqrt{n} = 3.09$, and the confidence interval is

$$[16.3 - 3.09, 16.3 + 3.09] = [13.21, 19.39].$$

b. How would the confidence interval widths change if n is increased to 100?

Since the width of the interval is $2z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}$, and when n is increased by a factor of 4 then $\sqrt{n}$ will be increased by a factor of 2, therefore, the width of the interval will be halved.

15. We want to estimate the mean output voltage of a batch of electrical power supply units. A random sample of 10 units is tested, and the sample mean is calculated to be 110.5 volts. Assume that the measurements are normally distributed with $\sigma = 3$ volts.

a. Calculate a two-sided 95% confidence interval on the mean output voltage. Suppose that the specifications on the true mean output voltage are 100 ± 2.5 volts. Explain how the confidence interval can be used to check whether these specifications are met.

The two-sided interval for confidence level 0.95 is

$$\left[\overline{X} - z_{0.025}\,\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{0.025}\,\frac{\sigma}{\sqrt{n}}\right].$$

With $n = 10$, $\sigma = 3$, and $\overline{X} = 110.5$, and since $z_{0.025} = 1.96$, this interval is

$$\left[\overline{X} - 1.96\,\frac{3}{\sqrt{10}}, \overline{X} + 1.96\,\frac{3}{\sqrt{10}}\right] = [108.6, 112.4].$$

This is a subinterval of 110 ± 2.5, so at the 95% confidence level, the specifications are met.

b. Suppose that only the lower specification is applicable, since the main concern is that too low a voltage may cause malfunction of the equipment. Calculate an appropriate one-sided 95% cofidence bound on the mean output voltage and explain how it can be used to check whether the lower specification limit is met.

The interval for confidence level $1 - \alpha$ is

$$\left[\overline{X} - z_{\alpha}\,\frac{\sigma}{\sqrt{n}}, \infty\right),$$

so this confidence interval is

$$[110.5 - 1.645\,\frac{3}{\sqrt{10}}, \infty) = [108.9, \infty).$$

Since 109.9 is greater than the lower limit 107.5, the specification is met.

Note how this one-sided specification is more lenient than the two-sided one.

Math 218 Home Page at
`http://math.clarku.edu/~djoyce/ma218/`