

Solutions

Answer 1. The formula is

$$z = \frac{w - 52}{2}.$$

Answer 2. $(2 + 6/1) + (2 + 6/2) + (2 + 6/3) = 8 + 5 + 4 = 17$.

Answer 3. The calculations are

$$\bar{x} = \frac{12 + 15 + 13 + 12}{4} = 13,$$

$$s^2 = \frac{1}{3}[(12 - 13)^2 + (15 - 13)^2 + (13 - 13)^2 + (12 - 13)^2] = 2,$$

$$s = \sqrt{2} = 1.4142.$$

The median is 12.5. Since the mean is larger than the median, we conclude that the sample is right-skewed.¹

Answer 4. The calculations are

$$E[X] = 0.6(10) + 0.3(20) + 0.1(30) = 15,$$

$$\text{Var}(X) = 0.6(10 - 15)^2 + 0.3(20 - 15)^2 + 0.1(30 - 15)^2 = 45,$$

$$\text{SD}(X) = \sqrt{45} = 6.7082.$$

Answer 5. We expect

$$E[\bar{X}] = 200, \quad \text{SD}(\bar{X}) = \frac{10}{\sqrt{25}} = 2.$$

This is the Central Limit Theorem at work. The distribution will be centered around the mean, and the standard deviation (in this case, also called the standard error of \bar{X}) will shrink as n gets bigger.

Answer 6a. Its skewness is pretty small, so yes. (Absolute value of 1 indicates at least mild skewness; anything less is unclear.)

¹Here's some intuition. Consider sample $\{13, 14, 15\}$. This has median and mean both 14, and it's also symmetrical. Now instead consider $\{13, 14, 100\}$. The median is still 14, but the mean is well above 14, and it's also right-skewed.

Answer 6b. 95% of observations would lie within

$$[\bar{x} \pm 2 \times s] = [0.6663 \pm 2 \times 0.1602] = [0.3459, 0.9867].$$

Answer 6c. For 193 degrees of freedom, the 90% confidence interval has critical value of 1.6528, giving (approximately)

$$\left[56257.22 \pm 1.6528 \times \frac{14535.76}{\sqrt{194}} \right] = [54535.27, 57979.17].$$

Answer 6d. mean earlycareer, level(90)

Answer 6e. $H_0 : \mu = 60,000$, $H_1 : \mu \neq 60,000$.

$$t = \frac{56,257.22 - 60,000}{14535.76/\sqrt{194}} = -3.5864.$$

This exceeds the critical value 1.6528 in magnitude, so we reject the null hypothesis.

Answer 6f. We fail to reject the null because the p -value is greater than our chosen significance level. In words, the t -statistic we find isn't quite improbable enough to reject based on our standard of 5% significance.

Answer 7. The interquartile range is given by

$$IQR \equiv Q_3 - Q_1 = 5.9934 - 2.0566 = 3.9368.$$

Answer 8. Skewness is not zero (absolute value of 1 indicates at least mild skewness) and kurtosis is not 3, so it does not appear to be normally distributed.

Answer 9: b. Having an estimator that varies a lot depending on our random sample is an undesirable property. Having an estimator that is usually pretty close to the true population value, on the other hand, is a desirable property.

Just to refresh your memory, an **unbiased** estimator is an estimator that we expect, on average, to equal the true population value. For example, if $X_i \sim (\mu, \sigma^2)$, then the sample mean

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

is an example of an unbiased estimator because $E[\bar{X}] = \mu$. This can be shown mathematically by utilizing two properties of the expectations operator. First, for two random variables X and Y , we have

$$E[X + Y] = E[X] + E[Y].$$

Furthermore, for any real number a , we have

$$E[aX] = aE[X].$$

These two properties constitute a *linear operator*. Accordingly, we can calculate

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{X_1 + \dots + X_n}{n}\right] \\ &= \frac{1}{n}E[X_1 + \dots + X_n] \\ &= \frac{1}{n}\left(E[X_1] + \dots + E[X_n]\right) \\ &= \frac{1}{n}\left(\underbrace{\mu + \dots + \mu}_{n \text{ times}}\right) \\ &= \frac{1}{n}(n\mu) \\ &= \mu. \end{aligned}$$

A **consistent** estimator is one that converges (in probability) to the true population value as n gets bigger. In other words, as the sample size gets bigger (we get more data), we expect the estimate to get closer and closer to the true population value. Sample mean \bar{X} is also consistent because $se(\bar{X}) = s/\sqrt{n} \rightarrow 0$ as $n \rightarrow \infty$.