# Unit 9 Assignment Solutions

1. (a) We first find z* such that $P(-z^* < Z < z^*) = 0.93$. The total area to the left of z* is

thus equal to 0.93 plus the area of the left tail, which is $(1 - 0.93)/2 = 0.035$. So we find z* such that $P(Z < z^*) = 0.93 + 0.035 = 0.965$ to be $z^* = 1.81$. The 93% confidence interval for $\mu$ is therefore

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 375 \pm 1.81\left(\frac{100}{\sqrt{25}}\right) = 375 \pm 36.20 = (338.80,\ 411.20)$$

(b) If we repeatedly took random samples of 25 students and calculated the interval in a similar manner, then 93% of such intervals would contain the true mean textbook cost for one semester.

(c) A 98% confidence interval for $\mu$ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 375 \pm 2.326\left(\frac{100}{\sqrt{25}}\right) = 375 \pm 46.52 = (328.48,\ 421.52)$$

(d) When we raise the confidence level, the length of the interval increases.

2. (a) We know that the sample mean $\bar{x}$ is at the center of the interval, so $\bar{x} = (40.973 + 43.027)/2 = 42$. The formula for the confidence interval is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

The upper confidence limit is equal to $\bar{x} + z^* \frac{\sigma}{\sqrt{n}} = 43.027$, so we solve for z*:

$$\bar{x} + z^* \frac{\sigma}{\sqrt{n}} = 43.027 \;\Rightarrow\; z^* = (43.027 - \bar{x})\left(\frac{\sqrt{n}}{\sigma}\right) = (43.027 - 42)\left(\frac{\sqrt{100}}{5}\right) = 2.054,$$

which is the value of z* for a 96% confidence interval.

(b) We know $\bar{x}$ is at the center of the interval, so $\bar{x} = (23.7 + 26.3)/2 = 25$. The formula for the confidence interval is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 25 \pm 1.645\left(\frac{\sigma}{\sqrt{n}}\right)$$

The upper confidence limit is equal to $\bar{x} + z^* \frac{\sigma}{\sqrt{n}} = 26.3$, so we solve for $\frac{\sigma}{\sqrt{n}}$:

$$\bar{x} + z^* \frac{\sigma}{\sqrt{n}} = 26.3 \;\Rightarrow\; \frac{\sigma}{\sqrt{n}} = \frac{26.3 - \bar{x}}{z^*} = \frac{26.3 - 25}{1.645} = 0.7903$$

The 99% confidence interval for $\mu$ is therefore

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 25 \pm 2.576\left(\frac{\sigma}{\sqrt{n}}\right) = 25 \pm 2.576(0.7903) = 25 \pm 2.0358 = (22.9642,\ 27.0358)$$

3. (a) $n = \left(\frac{z^* \sigma}{m}\right)^2 = \left(\frac{1.645(5.75)}{3}\right)^2 = 9.94 \approx 10$

Note that we must assume in this case that commuting distances follow a normal distribution, since the sample size is below 30.

(b) $n = \left(\frac{z^* \sigma}{m}\right)^2 = \left(\frac{1.645(5.75)}{1.5}\right)^2 = 39.76 \approx 40$

(c) For a given confidence level, we see that when we cut the desired margin of error in half, we require 40/10 = 4 times as many individuals in our sample.

(d) If we reduce the desired margin of error by a factor of k, while holding the confidence level constant, we require $k^2$ times as many individuals in our sample. Therefore, if we decrease the desired margin of error from 3 to 1 (i.e. decrease it by a factor of three), then we require $3^2 = 9$ times more individuals in our sample, or a sample of size $n = 9(10) \approx 90$

(e) $n = \left(\frac{z^* \sigma}{m}\right)^2 = \left(\frac{2.576(5.75)}{3}\right)^2 = 24.37 \approx 25$

(f) If we want to be more confident (i.e. 99% confident) that our interval contains the population mean, then we have to take a larger sample.

(g) It is the sample size, and not the population size, that affects the margin of error (provided that the size of the population is large compared to the size of the sample, which in this case it is). Therefore, even though Vancouver has five times as many people, the sample size required is the same as for Winnipeg, n = 10.

4. (a) Let $\alpha = 0.01$.

We are testing the hypotheses

$H_0$: The true mean voltage of all batteries on this production line is 240 V.
$H_a$: The true mean voltage of all batteries on this production line is less than 240 V.

Equivalently, $H_0$: $\mu = 240$ vs. $H_a$: $\mu < 240$.

We will reject $H_0$ if the P-value $\leq \alpha = 0.01$.

The test statistic is $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{238.3 - 240}{4.1 / \sqrt{16}} = -1.66$.

The P-value is $P(Z \leq -1.66) = 0.0485$.

Since the P-value $> \alpha = 0.01$, we fail to reject the null hypothesis.. At the 1% level of significance, we have insufficient evidence to conclude that the true mean voltage of all batteries produced on this production line is less than 240 V.

(b) If the true mean voltage of all batteries produced on this production line was 240 V, then the probability of selecting a sample with a mean at least as low as 238.3 V would be 0.0485.

5. (a) Even though the distribution of breaking strengths is not normal, the Central Limit Theorem tells us that, regardless of the form of the population distribution, the sampling distribution of $\overline{X}$ will be approximately normal, provided that the sample size is high. Our sample size here is 40, which is sufficiently high for the Central Limit Theorem to apply.

(b) Let $\alpha = 0.10$.

We are testing the hypotheses

$H_0$: The true mean breaking strength of this type of plastic is 7.0.
$H_a$: The true mean breaking strength of this type of plastic is greater than 7.0.

Equivalently, $H_0$: $\mu = 7.0$ vs. $H_a$: $\mu > 7.0$.

We will reject $H_0$ if the P-value $\leq \alpha = 0.10$.

The test statistic is $z = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{7.1 - 7.0}{0.25 / \sqrt{40}} = 2.53$..

The P-value is $P(Z \geq 2.53) = 1 - P(Z < 2.53) = 1 - 0.9943 = 0.0057$.\

Since the P-value $= 0.0057 < \alpha = 0.10$, we reject the null hypothesis. At the 10% level of significance, we have sufficient evidence to conclude that the true mean breaking strength of this type of plastic is greater than 7.0.

6. (a) In this question, we are given both the population standard deviation $\sigma$ and the sample standard deviation s. Since we know $\sigma$, we don't need to estimate it, and we use the z distribution for our inference. A 95% confidence interval for the true mean GPA is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 3.20 \pm 1.960\left(\frac{0.58}{\sqrt{50}}\right) = 3.20 \pm 0.16 = (3.04,\ 3.36)$$

(b) If we took repeated samples of 50 students from this university and calculated the interval in a similar manner, 95% of all such intervals would contain the true mean GPA.

(c) Let $\alpha = 0.05$.

We are testing the hypotheses

$H_0$: The true mean GPA of all students at this university is 3.00.
$H_a$: The true mean GPA of all students at this university differs from 3.00.

Equivalently, $H_0$: $\mu = 3.00$ vs. $H_a$: $\mu \neq 3.00$.

We will reject $H_0$ if the P-value $\leq \alpha = 0.05$.

The test statistic is $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{3.20 - 3.00}{0.58 / \sqrt{50}} = 2.44.$.

The P-value is $2P(Z > 2.44) = 2(1 - P(Z < 2.44)) = 2(1 - 0.9927) = 2(0.0073) = 0.0146$.

Since the P-value $< \alpha = 0.05$, we reject the null hypothesis. At the 5% level of significance, we have sufficient evidence to conclude that the true mean GPA at this university differs from 3.00.

(d) If the true mean GPA of all students at this university was 3.00, the probability of selecting a sample of 50 students with a mean GPA at least as extreme as 3.20 would be 0.0146.

(e) Since this is a two-sided test, and since the confidence level (95%) and the level of significance (5%) add up to 100%, we can use the interval in (a) to conduct the test. Since the value $\mu_0 = 3.00$ is not contained in the 95% confidence interval for $\mu$, we reject $H_0$.

7. (a) A 95% confidence interval for the true mean radioactivity is:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 4.75 \pm 2.045\left(\frac{0.87}{\sqrt{30}}\right) = 4.75 \pm 0.32 = (4.43,\ 5.07)$$

Note that we use the upper 0.025 critical value $t^* = 2.045$ from the t distribution with $n - 1 = 29$ degrees of freedom.

(b) Let $\alpha = 0.05$.

We are testing the hypotheses

$H_0$: The true mean radioactivity for the city's drinking water is 5 pCi/L.
$H_a$: The true mean radioactivity for the city's drinking water is less than 5 pCi/L.

Equivalently, $H_0$: $\mu = 5$ vs. Ha: $\mu < 5$.

We will reject $H_0$ if the P-value $\leq \alpha = 0.01$.

The test statistic is $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{4.75 - 5.00}{0.87/\sqrt{30}} = -1.57$.

The P-value is $P(T(29) \leq -1.57) = P(T(29) \geq 1.57)$ by the symmetry of the t distributions. We see from the t table that $P(T(29) \geq 1.311) = 0.10$ and $P(T(29) \geq 1.699) = 0.05$. Since $1.311 < 1.57 < 1.699$, it follows that our P-value is between 0.05 and 0.10.

Since the P-value $> \alpha = 0.05$, we fail to reject $H_0$. At the 5% level of significance, We have insufficient evidence to conclude that the true mean radioactivity level for the city's drinking water is less than 5 pCi/L.

(c) If the true mean radioactivity was 5 pCi/L, then the probability of selecting a sample of 30 specimens with a mean radioactivity at least as low as 4.75 pCi/L would be between 0.05 and 0.10.

(d) No, we could not use the confidence interval to conduct the test because this is a one-sided test. Confidence intervals can only be used for two-sided tests.

8. (a) If we want very strong evidence before we are prepared to reject the null hypothesis, then we will only reject if $\bar{x}$ is very high, which means the P-value must be low. We reject $H_0$ when the P-value $\le \alpha$, so if we want strong evidence before we are willing to reject $H_0$, we should use a **low** level of significance.

(b) Let $\alpha = 0.01$.

We are testing the hypotheses

$H_0$: The true mean strength of spruce is 3500 psi.
$H_a$: The true mean strength of spruce is greater than 3500 psi.

Equivalently, $H_0$: $\mu = 3500$ vs. $H_a$: $\mu > 3500$

We will reject $H_0$ if the P-value $\le \alpha = 0.01$.

The test statistic is $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{3600 - 3500}{225/\sqrt{25}} = 2.22$.

The P-value is $P(T(24) \ge 2.22)$. We see from the t table that $P(T(24) \ge 2.172) = 0.02$ and $P(T(24) \ge 2.492) = 0.01$. Since $2.172 < t = 2.22 < 2.492$, our P-value is between 0.01 and 0.02.

Since the P-value $> \alpha = 0.01$, we fail to reject the null hypothesis. At the 1% level of significance, we have insufficient evidence to conclude that the true mean strength of spruce is greater than 3500 psi.

(c) If the true mean strength of Southern Pine was 3500 psi, then the probability of getting a sample mean at least as high as 3600 would be between 0.01 and 0.02.

9. (a) A 90% confidence interval for the true mean resistance of all wires of this type is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 197 \pm 1.721\left(\frac{10}{\sqrt{22}}\right) = 197 \pm 3.67 = (193.33,\ 200.67)$$

where $t^* = 1.721$ is the upper 0.05 critical value from the t distribution with $n - 1 = 21$ degrees of freedom.

(b) If we were to repeatedly take samples of 22 wires and construct a confidence interval in a similar manner, then 90% of such intervals would contain the true mean resistance of all wires of this type.

(c) Let $\alpha = 0.10$.

We are testing the hypotheses

$H_0$: The true mean resistance of all wires of this type is 200 Ohms.
$H_a$: The true mean resistance of all wires of this type differs from 200 Ohms.

Equivalently, $H_0$: $\mu = 200$ vs. $H_a$: $\mu \neq 200$.

We will reject $H_0$ if the P-value $\leq \alpha = 0.10$.

The test statistic is $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{197 - 200}{10/\sqrt{22}} = -1.41.$.

The P-value is $2P(T(21) \leq -1.41) = 2P(T(21) \geq 1.41)$ by the symmetry of the t distributions. We see from the t table that $P(T(21) \geq 1.323) = 0.10$ and $P(T(21) \geq 1.721) = 0.05$. Since $1.323 < 1.41 < 1.721$, it follows that $P(T(21) \geq 1.41)$ is between 0.05 and 0.10. Our P-value is double this probability, so the P-value of the test is between 2(0.05) and 2(0.10), i.e. between 0.10 and 0.20.

Since the P-value $> \alpha = 0.10$, we fail to reject the null hypothesis. At the 10% level of significance, we have insufficient evidence to conclude that the true mean resistance of all wires of this type differs from 200 Ohms.

(d) Since this is a two-sided test, and since the confidence level (90%) and the level of significance (10%) add up to 100%, we can use the interval in (a) to conduct the test. Since the value $\mu_0 = 200$ is contained in the 90% confidence interval for $\mu$, we fail to reject $H_0$.