HW5 Solutions

1  IPS 6.20 [3 points]

We have $\bar{x} = 33.4$, $\sigma = 19.6$ and $n = 31$. Using our standard formula for a CI for $\bar{X}$, $\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$, we have $33.4 \pm 1.96 \frac{19.6}{\sqrt{31}}$, giving an interval:

$$(26.50, 40.30)$$

3 points for correct interval (using $Z = 2$ will give $(26.36, 40.44)$ for full marks); if wrong, 1 point for correct formula, 1 point for correct $Z$ score (1.96 or 2 are fine).

2  IPS 6.24 [2 points]

We have $\bar{x} = 1050$, $\sigma = 220$ and $n = 10$. Using our standard formula for a CI for $\bar{X}$, $\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$, we have $1050 \pm 1.96 \frac{220}{\sqrt{10}}$, giving an interval:

$$(913.6, 1186.4)$$

2 points for correct interval (using $Z = 2$ will give $(910.9, 1189.1)$ for full marks); if wrong, 1 point for either correct formula use or correct $Z$ score (1.96 or 2 are fine) (correct formula and correct $Z$ score but wrong interval is still just one point).
3 IPS 6.36 [4 points]

(a) We have $\bar{x} = 10.0023$, $\sigma = 0.0002$ and $n = 5$. Using our standard formula for a CI for $\bar{X}$, $\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$, we have $10.0023 \pm 2.33 \frac{0.0002}{\sqrt{5}}$, giving an interval:

$$(10.00209, 10.00251)$$

Recall that for a 98% interval, we need to leave 1% in each tail. Looking at our table (or elsewhere, e.g. R), we find $Z = 2.33$.

(b) As we can see from our formula, our margin of error is $\pm Z \frac{\sigma}{\sqrt{n}}$. Thus we need $Z \frac{\sigma}{\sqrt{n}} = 0.0001$. We want 98% confidence, so again we have $Z = 2.33$. $\sigma$ is fixed at 0.0002, so we’re just solving for $n$. This gives $n = \left(\frac{2.33 \times 0.0002}{0.0001}\right)^2 = 21.65$. This means that we’ll have a margin of error larger than 0.0001 for $n \leq 21$ and smaller than 0.0001 for $n \geq 22$. In these cases, we usually implicitly assume that we need a margin of error as small as asked for or smaller, so the correct answer would be that we need at least 22 measurements. You can also verify that $n = 22$ gives a margin closer to 0.0001 than $n = 21$.

2 points for (a): 2 for correct interval; if wrong, 1 point for either correct $Z$ score or correct formula use. 2 points for (b): full marks also given if solution arrived at by simply trying different values of $n$ until correct margin of error found. $n = 21$ is also OK if some justification given, 1 mark lost if no indication given at all as to how $n$ was found.

4 [3 points]

(a) For a 95% interval, we use $Z = 1.96$. Our interval is $\bar{x} \pm Z \frac{\sigma}{\sqrt{n}} = 143 \pm 1.96 \frac{24}{\sqrt{86}}$, which gives $(137.84, 148.16)$. 

2
(b) It gets smaller - if we’re happy to have less confidence in our interval, we don’t need it to be as long. Actual interval not required, but it’s (138.67, 147.33).

(c) It gets larger: in order to be more confident that we trapped the true mean in our interval, we’ll need it to be longer. Again you don’t have to give the interval, but it’s (136.22, 149.78) if you really want it.

1 point for (a): correct interval required for (using $Z = 2$ to get (137.01, 148.99) is fine too). For (b) and (c), to get both points, at least one of the two must contain some sort of explanation - either the confidence change, or give the new $Z$ value, or give the interval; else at most one point for the two parts.

5 IPS 6.62 [3 points]

“Significantly different at the 0.01 level” means that some statistic(s) (i.e. number(s), we usually call them test statistics) that measures the grave goods (e.g. estimated value of goods and artifacts, or total number of goods and artifacts) differs enough between the pig skull and the non-pig skull burials that the probability of seeing such a difference, or an even greater difference, is $\leq 0.01$ (or 1%) if there really was no true difference between the sites.

In other words, if pig skulls signified nothing at all, our statistic for measuring goods and artifacts would only differ by random chance between these types of sites, and the probability of this random chance producing an even more unlikely difference than we actually see is $\leq 0.01$, thus we say that the difference is significant at the 0.01 level.

This means that the implicit null hypothesis of “no difference” (i.e. that the pig skulls signify nothing) is either not true, or we must have been very lucky
(or unlucky, depending on your perspective and distaste for pig skulls) to see such a connection between the skulls and the good and artifacts in the sites - lucky in the sense that we should only see such a difference less than one in every one hundred such tests.

3 points in total: for full marks, a reference to what the 0.01 literally means is required (something like an implicit statement about \( p \)-values). Any one of the above three paragraphs would be good enough alone (although the last one most directly addresses the question). A vaguer statement, such as “the data is unlikely if no difference existed” that at least addresses the question, is worth 2 points. 1 point for a decent attempt that misses the mark.

6 IPS 6.70 [2 points]

Our \( Z \)-score is formed the usual way:

\[
Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{10.2 - 8.9}{2.5/\sqrt{6}} = 1.274
\]

The \( p \)-value is the probability, if the null is true, of seeing data at least as extreme or more extreme than what we actually saw. In our case, since we have a one-sided test (due to our one-sided alternative), more extreme data would be to see an even higher average of new words, i.e. seeing \( \bar{X} \geq 10.2 \). So we want \( P(\bar{X} \geq 10.2) \), assuming \( \mu = 8.9 \). We know the \( Z \)-score, so this is the same as \( P(Z \geq 1.274) \). Checking our tables (or R etc), this gives \( p = 0.101 \). From here, we conclude that we can’t reject the null, since this is not a particularly small \( p \)-value. Generally if not told otherwise, we assume we’re comparing to \( \alpha = 0.05 \), and here, \( 0.101 > 0.05 \). But even without this, we rarely would use any \( \alpha \) value larger than 0.1, and \( 0.101 > 0.1 \). Finally, we can just look at the \( p \)-value itself and see that it’s not tiny without an explicit rule of thumb to compare it to... This of course doesn’t mean that we
believe these sonnets are by our poet, only that we can’t say with confidence that they belong to someone else.

2 points for correctly calculating the $p$-value and concluding that we can’t reject the null hypothesis. Also full marks for calculating the correct $p$-value and then giving good reasoning for not being sure if it’s small enough to reject, due to $\alpha$ not being given in the question (i.e. the value to compare $\alpha$ to). If not: 1 point for correct $p$-value or 1 point for making a correct conclusion relative to whatever $p$-value you got.

Note: if you got a $p$-value of about 0.2, you probably calculated the two-sided $p$-value by accident.

7 IPS 6.72 [3 points]

(a) Let $\mu$ be the mean caloric intake for the Canadian women. We have:

\[ H_0 : \mu = 2811.5 \]
\[ H_a : \mu < 2811.5 \]

Note that the test is one-sided, since the question is asking for evidence of deficient intake. If your null is $H_0 : \mu \geq 2811.5$, that’s fine too.

(b) In this case, data as extreme or more extreme than what we saw would be an average of 2403.7 or less (one way to think about this: imagine we collected data for a new set of 201 women). So we want $P(\bar{X} \leq 2403.7)$, assuming we have $\mu = 2811.5$. We have $\bar{x} = 2403.7$, $\sigma = 880$ and $n = 201$ (not 324). Our $Z$-score is then formed the usual way:

\[ z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2403.7 - 2811.5}{880/\sqrt{201}} = -6.57 \]
So, $P(\bar{X} \leq 2403.7) = P(Z \leq -6.57)$. If you really want to know this value, it’s about $2.5 \times 10^{-11}$ (so, tiny then). This $z$-score isn’t on your tables, so it’s fine to say that this is basically zero; it’s also fine to conclude that this value must be $< 0.003$, which is the smallest value on your table. Anyway, whichever of these three methods you go with, we reject our null of sufficient intake ($\mu = 2811.5$), and conclude that $\mu < 2811.5$. In plain English, we have (strong) evidence that these women are not consuming enough calories.

1 point for (a). 1 point for correct $p$-value in (b) and 1 point for plain English interpretation (interpretation is relative to the $p$-value, so if the $p$-value is wrong, the interpretation of it can still be correct).

8 IPS 6.124 [4 points]

(a) Ideally, we all found the formula that gives $se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for the standard deviation. Thus we can construct 95% intervals, using the formula $\hat{p} \pm Zse(\hat{p})$. We want 95%, so we use $Z = 1.96$ (again, 2 is fine). This gives:
<table>
<thead>
<tr>
<th>Occupation</th>
<th>( \hat{p} )</th>
<th>n</th>
<th>CI:low</th>
<th>CI:high</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Professional</td>
<td>0.23</td>
<td>2447</td>
<td>0.213</td>
<td>0.247</td>
</tr>
<tr>
<td>2 Mangerial</td>
<td>0.22</td>
<td>2552</td>
<td>0.204</td>
<td>0.236</td>
</tr>
<tr>
<td>3 Administrative</td>
<td>0.17</td>
<td>2309</td>
<td>0.155</td>
<td>0.185</td>
</tr>
<tr>
<td>4 Sales</td>
<td>0.15</td>
<td>1811</td>
<td>0.134</td>
<td>0.166</td>
</tr>
<tr>
<td>5 Mechanical</td>
<td>0.12</td>
<td>1979</td>
<td>0.106</td>
<td>0.134</td>
</tr>
<tr>
<td>6 Service</td>
<td>0.13</td>
<td>2592</td>
<td>0.117</td>
<td>0.143</td>
</tr>
<tr>
<td>7 Operator</td>
<td>0.12</td>
<td>2782</td>
<td>0.108</td>
<td>0.132</td>
</tr>
<tr>
<td>8 Farm</td>
<td>0.08</td>
<td>571</td>
<td>0.058</td>
<td>0.102</td>
</tr>
</tbody>
</table>

(b) At this time, we probably won’t fully get into the discussion about how to interpret confidence interval overlap or lack thereof. Turns out that if two intervals have no overlap, we know that the means are indeed significantly different, but it’s possible to have some overlap in the intervals and yet the appropriate two-sample test would show a significant difference. Visually, we might say that farmers are chillin’, then operator / service / mechanical are in a similar group slightly higher, then maybe sales and admin aren’t that different, and finally managerial and professional are close, being the worst.

Any summary that makes some sense is fine here, no need to have correctly identified groups.

(c) Your friend might be concerned with one of two things. Firstly, the concern your friend might have is that you don’t really have \( n \) independent samples in each category: you ask people the same question more than once, you try to ask them every two years, and so you don’t necessarily have independent samples. It’s possible that asking someone their stress level every two years will give an independent response, but it’s likely that you have some dependence, and the formula assumes
independence. So you should probably have a larger standard deviation
and a larger confidence interval (or in other words, your intervals will
probably have less than 95% coverage).

Secondly, that the formula should be $\sqrt{\frac{np(1-p)}{n}}$, i.e. with the true $p$, but
we’re using the estimated $p$, i.e. $\hat{p}$ in the formula instead, since we don’t
have the true $p$ (if we did, we wouldn’t need to bother making a CI at
all). Due to the large values of $n$ that we have, this isn’t such a huge
problem, so your friend can chill. Note: since this question specifically
refers to the use of the standard deviation formula, the answer isn’t
something to do with normality assumptions, and thus central limit
theorem isn’t going to make your friend relax.

2 points for (a): both if all intervals are correct, one if not but something
correct is done (e.g. correct formula etc) [If nearly all intervals are correct,
you can let minor numerical errors slide]. 1 point for (b), details at the end
of (b), not too strict. 1 for (c), either answer gets the point.

9 IPS 6.128 [3 points]

Let’s start with $n = 10$. We have $\bar{x} = 4$, $n = 10$, and a null hypothesis $\mu = 0$.
For the $p$-value, we’re looking for data as or more extreme than what we got:
this would be values of $\bar{X}$ larger than 4, thus we want $P(\bar{X} \geq 4)$. To compute
this, we do the usual, and get our test statistic, the $Z$-score:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{4 - 0}{14/\sqrt{10}} = 0.904$$

Thus, the $p$-value is $P(Z \geq 0.904) = 0.183$. And if we want: since $0.183 > \alpha = 0.05$, we don’t reject at the 0.05 level.

Firstly, the table of values:
<table>
<thead>
<tr>
<th></th>
<th>Test Statistic (Z)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.90</td>
<td>0.183</td>
</tr>
<tr>
<td>2</td>
<td>1.28</td>
<td>0.101</td>
</tr>
<tr>
<td>3</td>
<td>1.56</td>
<td>0.059</td>
</tr>
<tr>
<td>4</td>
<td>1.81</td>
<td>0.035</td>
</tr>
<tr>
<td>5</td>
<td>2.02</td>
<td>0.022</td>
</tr>
</tbody>
</table>

And plotted:
I’ve added the 0.05 significance line above, you don’t need it.

Any reasonable summary: something along the lines of the larger the sample, the easier is it to prove significance (but only easier if there is a true difference! You don’t need to add that part).

1 point for correct $Z$ and $p$ values, 1 for correct graphs, 1 for some sort of sensible conclusion.

10 [3 points]

(a) The sample standard deviation is $s = 24.91$, and we’ll assume that $\sigma$ takes the same value. We have $\bar{x} = 35.36$ from the data, and $z = 1.96$ because we want a 95% interval, and of course $n = 29$. The usual
formula gives:

\[(26.30, 44.43)\]

(b) Firstly, we already know we’re going to reject 25 at the two-sided 0.05 level, since it’s not in our 95% CI given above (we can’t always make the same conclusion about one-sided tests: CIs naturally pair with two-sided tests). However, let’s go through the mechanics of the hypothesis test - and we have to, since we’re asked for it all specifically. We have:

\[
H_0 : \mu = 25 \\
H_1 : \mu \neq 25
\]

where \(\mu\) is of course the mean cost of playing 18 holes. We have \(\bar{x} = 35.36\), with a z-score (test statistic) of \(z = \frac{35.36 - 25}{24.91/\sqrt{29}} = 2.24\). Since the test is two-sided, a more extreme test statistic would be a z value above 2.24 or below −2.24. We can also write this as \(|Z| \geq 2.24\) if we like. \(P(Z \geq 2.24) = 0.0125\), and so our two-sided test gives a \(p\)-value of 0.025 (we can just double the above probability, since we’re adding \(P(Z \leq -2.24)\), and this is the same probability). 0.025 < 0.05, therefore we reject our hypothesis at the 0.05 level, and conclude that the mean green fee is above $25.

1 point for (a), correct interval. 1 point for correctly stating the hypotheses, 1 point for correct \(p\)-value and conclusion.