SUTD ESD

Statistics 2017

Assignment 2 Solutions

Question 1. (a) $H_0 : \mu = 25$ vs $H_1 : \mu < 25$ (μ is the mean commuting time in minutes).

(b) $H_0 : \mu = 3.4$ vs $H_1 : \mu > 3.4$ (μ is the mean fat content in grams per cup).

Question 2. (a) We have $\bar{x} = 16.3$, $\alpha = 0.1$ and $n = 100$; the CI is given by

$$\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = [15.313, 17.287].$$

(b) The CI is given by

$$\left[\bar{x} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}\right] = [15.304, 17.296].$$

(c) Upper 90% CI:

$$\left(-\infty,\ \bar{x} + t_{n-1,1-\alpha}\frac{s}{\sqrt{n}}\right] = (-\infty,\ 17.074].$$

Lower 90% CI:

$$\left[\bar{x} - t_{n-1,1-\alpha}\frac{s}{\sqrt{n}},\ \infty\right) = [15.526,\ \infty).$$

(d) The wider confidence interval reflects the greater uncertainty in (b): the true value of σ is unknown and is estimated using s. (Mathematically, this is seen in the fact that a t-distribution has heavier tails than the standard normal distribution.)

Question 3. (a) By the central limit theorem, U is approximately normal with mean 40 and standard deviation $15/\sqrt{50}$; V is approximately normal with mean 40 and standard deviation $15/\sqrt{100}$.

(b)

$$P(38 \le U \le 42) = P\left(|Z| \le \frac{2}{15/\sqrt{50}}\right) = 0.6542; \qquad P(38 \le V \le 42) = P\left(|Z| \le \frac{2}{15/\sqrt{100}}\right) = 0.8176.$$

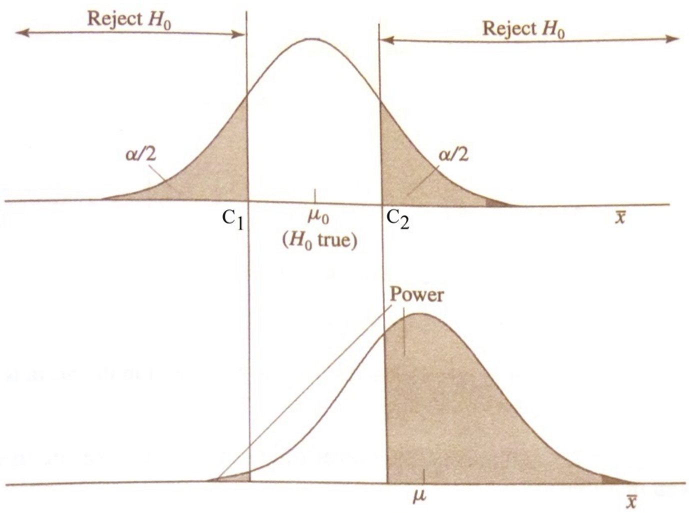
Question 4. (a) To compute the p-value, we first assume that H_0 is true, that is, $\mu = 0$.
Then for (i), the p-value is

$$P\left(|Z| \ge \frac{0.1 - 0}{1/\sqrt{100}}\right) = P(|Z| \ge 1) = 0.3173;$$

for (ii), $P(|Z| \ge 2) = 0.0455$; for (iii), $P(|Z| \ge 3) = 0.0027$.
Note that the p-values are two-sided, since the alternative hypothesis is two-sided.

(b) For (i) and (ii), the p-value is greater than α, so we do not reject H_0; for (iii), the p-value is smaller than α, so we reject H_0.

Question 5. (a) The power is the probability of rejecting H_0 when H_1 is true. Since the alternative is two-sided, there are two rejection regions, shaded under the top bell curve.



The boundaries of these regions are at $c_1 = \mu_0 - z_{1-\alpha/2}\,\sigma/\sqrt{n}$ and $c_2 = \mu_0 + z_{1-\alpha/2}\,\sigma/\sqrt{n}$.

Now if H_1 is true and the mean is at some point $\mu \neq \mu_0$, then the power is the probability shaded under the bottom bell curve. (We chose to depict μ on the right of μ_0, but this does not change the generality of our calculations.)

The desired probability is given by the sum of the two shaded parts:

$$\begin{aligned}
P(\bar{x} \geq c_2) + P(\bar{x} \leq c_1) &= P\left(Z \geq \frac{c_2-\mu}{\sigma/\sqrt{n}}\right) + P\left(Z \leq \frac{c_1-\mu}{\sigma/\sqrt{n}}\right) \\
&= 1 - \Phi\left(\frac{c_2-\mu}{\sigma/\sqrt{n}}\right) + \Phi\left(\frac{c_1-\mu}{\sigma/\sqrt{n}}\right) \\
&= \Phi\left(\frac{\mu-c_2}{\sigma/\sqrt{n}}\right) + \Phi\left(\frac{c_1-\mu}{\sigma/\sqrt{n}}\right) \qquad \big(\text{as } 1-\Phi(x) = \Phi(-x)\big) \\
&= \Phi\Big(\frac{(\mu-\mu_0)\sqrt{n}}{\sigma} - z_{1-\alpha/2}\Big) + \Phi\Big(\frac{(\mu_0-\mu)\sqrt{n}}{\sigma} - z_{1-\alpha/2}\Big),
\end{aligned}$$

as required.

(b) If $\mu_0 - \mu$ is negative and large (in absolute value), then second term above becomes negligible (since the argument of Φ becomes small and negative) – this can be seen in the diagram, where the bottom left shaded area is very small. Likewise, if $\mu_0 - \mu$ is positive and large, then the first term above becomes negligible.

In either case, if we omit the negligible term and equate the other term to $1-\beta$, then after some algebra, we find that

$$n = \left(\frac{(z_{1-\alpha/2} + z_{1-\beta})\sigma}{\mu - \mu_0}\right)^2.$$

Question 6. $H_0 : \sigma = 3500$, $H_1 : \sigma > 3500$.

Assume H_0 is true, then $\dfrac{(n-1)s^2}{\sigma^2}$ follows a χ^2 distribution with 24 degrees of freedom. The p-value, that is, the probability of obtaining $s \geq 4569$, can be computed using the *Excel* command

```
1 - chisq.dist(24*4569^2/3500^2, 24, 1).
```

The p-value is $0.01708 < 0.05$, so we reject the null hypothesis and conclude that the standard deviation is indeed greater than the claimed value.

Alternatively, we could construct the lower 95% one-sided CI for σ, which is $[3709.3, \infty)$. Since 3500 is outside this interval, we reject H_0.

Question 7. (a) $H_0 : \mu = 10$, $H_1 : \mu < 10$. (It is also fine to use $\geq$ in H_0.)

(b) The p-value is $P(\bar{X} \leq 8.7) = P\left(Z \leq \dfrac{8.7 - 10}{2/\sqrt{15}}\right) = 0.00591 < 0.05$, so there is statistically significant evidence to reject H_0.

(c) We need to compute the power of the test. From *exam practice* Q13, the correct formula to use here is

$$1 - \beta = \Phi\left(\frac{(\mu_0 - \mu)\sqrt{n}}{\sigma} - z_{1-\alpha}\right).$$

We are given $\mu_0 - \mu = 1.5$, $n = 15$, $\sigma = 2$ and $\alpha = 0.05$. Substituting these values into the formula, we obtain a power of 0.896.

Question 8. Note that we are testing whether *two* populations have the same variance (so the F distribution is required); this problem cannot be reduced to testing the variance of a single population, since we do not know the true variance for either population.

We have $H_0 : \sigma_1 = \sigma_2$, $H_1 : \sigma_1 > \sigma_2$, or equivalently, $H_0 : \sigma_1^2/\sigma_2^2 = 1$, $H_1 : \sigma_1^2/\sigma_2^2 > 1$.

Assuming normality, then we know that $\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$ is an $F_{8,8}$ random variable, therefore

$$P\left(\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \leq f_{8,8,0.95}\right) = 0.95.$$

Rearranging, and using $s_1 = 2.3$, $s_2 = 1.1$, we find that the lower 95% CI for σ_1^2/σ_2^2 is $[1.27, \infty)$. Since this interval does not contain the value 1, we reject H_0 and conclude that the new oven indeed provides more even heating.

Question 9. Because she always picks exactly 6 of the cups as tea-first, and the rest as milk-first, if she identifies k of the tea-first cups correctly, then she will automatically get k of the milk-first cups right. (Use a diagram if you are not convinced!)

So, to do at least as well as her means to either (i) correctly identify 5 of the tea-first cups (and hence also 5 of the milk-first cups); or (ii) correctly identify 6 of the tea-first cups (and hence also 6 of the milk-first cups) – that is, identify everything correctly.

There are $\binom{12}{6} = 924$ ways to pick 6 cups from the 12 cups (ignoring the order in which they are chosen). The probability of getting (ii) by random guessing is just $1/924$. The probability of (i) is $\binom{6}{5}\binom{6}{1}/924$, as the $\binom{6}{5}$ accounts for the number of ways to get 5 of the 6 tea-first cups right, and the $\binom{6}{1}$ accounts for the number of ways to get 1 of the tea-first cups *wrong*.

Adding up the two probabilities, we find that the p-value is $37/924 \approx 0.040 < \alpha$, therefore we conclude that there is significant evidence in support of her claim.