Homework 3 Solution

Statistical Inference, Spring 2021

1-

a. True.

b. False, P(Type 1 error) = P(reject H_0 | H_0 true) = α.

c. False, there is probability of Type 2 error.

d. True.

e. True.

f. False.As the sample size gets small, the z value decreases therefore we will less likely to reject the null hypothesis; more likely to fail to reject the null hypothesis, thus the power of the test decreases.

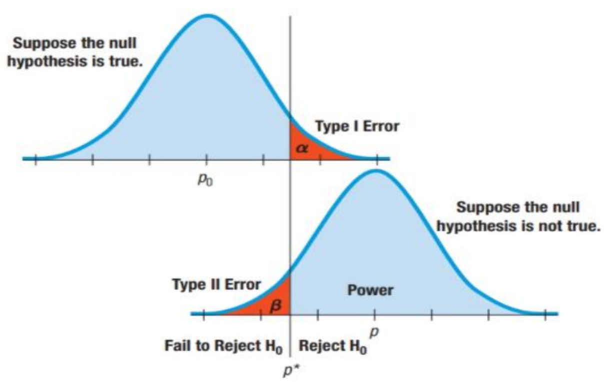
g. False. If all other things are held constant, then as α (significance level) decreases, so does the power of the test. This is because a smaller α means a smaller rejection region for the test and thus a lower probability of rejecting the null hypothesis. That translates to a less powerful test.

h. False. α =P (reject H_0 | H_0 true)

i. False. Power= P (reject H_0 | H_A true)

j. False. A Type I error occurs when the test statistic falls in the rejection region of the test and H_0 is true

k. False. As you can see below, if the difference between p and p_0 increases, the power will increase.



l. False. If the standard deviation is decreased, the SE and β will decrease so the power will increase.

2-

a. A type I error in terms of this problem will be if we decide that the supplier gives less mean weight of potato chips when actually he gives the correct amount.

b. A type II error in terms of this problem will be if we decide that the supplier gives the correct mean amount of potato chips when actually he gives less.

c. The customers view as more serious a type II error because no action will be taken by the supplier.

d. The supplier views as more serious a type I error because unnecessary action will be taken.

3- We are given the following information: $n = 50$, $\bar{x} = 115$, and $\sigma = 10$.

a. Construct a 97% confidence interval for the population μ.

Check conditions:

1. Independence: random & $n = 50 < 10\%$ of population → independence

2.Sample size/skew: $n = 100 \geq 30$ & sample not skewed→ nearly normal sampling distribution

$1 - \alpha = 0.97 \rightarrow z_{\frac{\alpha}{2}} = 2.17$. The confidence interval will be:

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$115 - 2.17 \frac{10}{\sqrt{50}} \leq \mu \leq 115 + 2.17 \frac{10}{\sqrt{50}}$$

$$111.931 \leq \mu \leq 118.069$$

b. We are 97% confident that μ falls in the above interval. Also, you can say that if all possible samples of the same size (here $n = 50$), were taken, 97% of them would include the true population mean μ, and only 3% would not.

c.

I. True.

II. False. 95% of the factory's batteries' **average lifetime** is in range of 112.23 to 117.77 hours.

III. False. The factory's batteries' true average lifetime **is** in the CI **or is not**.

IV. False. The factory’s batteries are the **population**, not any batteries in the market.

4-

a. $\hat{p} = \frac{4781}{100000} = 0.04781$

Check Conditions:

1- The size of sample is no more than 5% of the size of population it was drawn from.

2- $np(1-p) \geq 10$
$(100000)(0.04781)(1-0.04781) \geq 10$
$4552.42 \geq 10$
approximately normal distribution.

90% CI: $\hat{p} \pm z_{\propto/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, $z_{0.1/2} = 1.65$

$$0.04781 - 1.65\sqrt{\frac{0.04781(1-0.04781)}{100000}} = 0.04892$$

$$0.04781 + 1.65\sqrt{\frac{0.04781(1-0.04781)}{100000}} = 0.04669$$

90% CI: $(0.04669\,, 0.04892)$

b. $H_0 : p = 0.05$

$H_A : p \neq 0.05$

No. The H_0 is not within 90% CI, therefore 5% is not plausible.

5- (**R**)

6- $n = 50\,,\ \bar{y} = 25.9\ ,\ s = 5.6$

$H_0\!: \mu = 28$

$H_A\!: \mu < 28$

Check conditions:

1. Independence: random & $n = 50 < 10\%$ of population → independence

2.Sample size/skew: $n = 50 \geq 30$ & sample not skewed→ nearly normal sampling distribution

$$\bar{y} \sim N(mean = \mu = 28, SE = \frac{s}{\sqrt{n}} = \frac{5.6}{\sqrt{50}} = 0.792)$$

a. $p-value = P(\bar{y} < 25.9 | H_0: \mu = 28) = P(Z < -2.652) = 0.004 < 0.05 \rightarrow$ reject H_0; the data provide convincing evidence for H_A.

Test statistics: $Z = \frac{25.9 - 28}{0.792} = -2.652$

Based on sample information we can reject H_0 in favor of H_A.

b.

$$\mu_a = 27 \rightarrow \bar{y} \sim N(\mu = 27, SE = \frac{s}{\sqrt{n}} = \frac{5.6}{\sqrt{50}} = 0.792)$$

$\alpha = 0.05$, one sided test $\rightarrow P(Z < z_\alpha) = 0.05 \rightarrow z_\alpha = -1.645$

$$\begin{aligned}
\text{Type 2 Error } = \beta &= P\left(\frac{\bar{y} - 28}{0.792} > -1.645 | \bar{y} \sim N(\mu = 27, SE = 0.792)\right) \\
&= P\left(\bar{y} > -1.645 * 0.792 + 28 = 26.697 \middle| \bar{y} \sim N(\mu = 27, SE = 0.792)\right) \\
&= P\left(Z > \frac{26.697 - 27}{0.792}\right) = P(Z > -0.383) = 1 - P(Z < -0.383) \\
&= 1 - 0.352 = 0.648
\end{aligned}$$

c. No, because type II error will be made if we fail to reject H_0 , but in part (a) the p-value is too small to reject H_0.

7- (**R**)

8-

a. $n = \frac{(80)^2(1.645+1.96)^2}{(525-550)^2} = 133.1 \rightarrow n = 134$

b. $n = 100,\ \bar{y} = 542\ ,\ s = 76$

$H_0: \mu = 525$

$H_A: \mu > 525$

Check conditions:

1. Independence: random & $n = 100 < 10\%$ of population $\rightarrow$ independence

2. Sample size/skew: $n = 100 \geq 30$ & sample not skewed $\rightarrow$ nearly normal sampling distribution

$$\bar{y} \sim N(mean = \mu = 525, SE = \frac{s}{\sqrt{n}} = \frac{76}{\sqrt{100}} = 7.6)$$

$p - value = P(\bar{y} > 542 | H_0: \mu = 525) = P(Z > 2.237) = 0.0125 < 0.05$

Test statistics: $Z = \frac{542-525}{7.6} = 2.237$

$P(Z > 2.237) = 1 - P(Z < 2.237) = 1 - 0.9875 = 0.0125$

$P(Z > 2.237) = P(Z < -2.237) = 0.0125$

→ reject H_0; the data provide convincing evidence for H_A. → there is sufficient evidence to conclude that the mean mathematics achievement level has been increased.

9- The proportion of people experienced midmorning fatigue among ones who ate breakfast is 298/3,055 = 0.0975 ($\widehat{P_2}$) and The proportion of people experienced midmorning fatigue among ones who skipped breakfast is 81/744 = 0.1089 ($\widehat{P_1}$).

$H_0: p_1 = p_2 \quad H_A: p_1 > p_2 \qquad \alpha = 0.05$

We must first check that the sample size is adequate. Specifically, we need to ensure that we have at least 5 successes and 5 failures in each comparison group. In this example, we have more than enough successes (cases who experienced midmorning fatigue) and failures (cases who didn't experience midmorning fatigue) in each comparison group. The sample size is more than adequate so the following formula can be used:

$$z = \frac{\widehat{P_1} - \widehat{P_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Computing the test statistic:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{81 + 298}{744 + 3055} = 0.0998$$

$$z = \frac{\widehat{P_1} - \widehat{P_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 0.927$$

Decision rule: Reject H_0 if Z > 1.65

Conclusion: We do not reject H_0 because $Z = 0.927 < 1.65$. We do not have statistically significant evidence at $\alpha = 0.05$ to show that prevalence of midmorning fatigue is significantly more among persons who skip breakfast than those who do not.