Biostatistics Homework 5

By 𝕃umi (张鹿鸣 12112618)

1. One-sided confidence interval 1

In this question, we are interested in one sided confidence intervals. Thus we want to find:

$$\mathrm{P}(\mu \le b) = 0.95$$

Since the normal distribution is symmetric, and our population luckily follows a normal distribution with a know standard deviation, the one sided confidence interval for the lower tail and the upper bond will be just symmetric around $\bar{X}$. The margin of error can be calculated with a z score.

```
qnorm(0.95)
```

```
## [1] 1.644854
```

We get from R that when $z = 1.645$, $\mathrm{P}(Z < z) = 0.95$. We also know that:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Thus the one sided confidence interval of the lower tail will be C.I. $= (-\infty, \bar{X} + 1.645\frac{\sigma}{\sqrt{n}}]$.

Similarly, the one sided confidence interval of the upper tail will be C.I. $= [\bar{X} - 1.645\frac{\sigma}{\sqrt{n}}, +\infty)$.

2. One-sided confidence interval 2

Different from question 1, although our population still follows a normal distribution, this time, we don't know anything about it's variance. Thus, we need to substitute the population variance with the sample variance. Our random variable $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ follows a T-distribution. We want to find:

$$\mathrm{P}(\mu \le b) = 0.95$$

The T-distribution is also symmetric, so the upper tail and the lower tail is symmetric around $\bar{X}$. The margin of error can be calculated from the t score.

If we want to calculate the confidence interval for the lower tail, the t score is $t_{0.05,v-1}$. We also know that:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Thus the confidence interval for the lower tail will be C.I. $= (-\infty, \bar{X} + t_{0.05,v-1}\frac{s}{\sqrt{n}}]$.

Similarly, the confidence interval for the upper tail will be C.I. $= [\bar{X} + t_{0.95,v-1}\frac{s}{\sqrt{n}}, +\infty) = [\bar{X} - t_{0.05,v-1}\frac{s}{\sqrt{n}}, +\infty)$.

3.

T

F

$[0.7421, 0.8229]$ (See 3 supplement)

F

T

3 supplement We know from the problem that this is a test for proportion. Thus the confidence interval is:

```
qnorm(0.975)
```

```
## [1] 1.959964
```

```
313/400
```

```
## [1] 0.7825
```

$$\text{C.I.} = [0.7825 - 1.96\sqrt{\frac{(1-0.7825)\times 0.7825}{400}}, 0.7825 + 1.96\sqrt{\frac{(1-0.7825)\times 0.7825}{400}}]$$

Which is $[0.7421, 0.8229]$.

4 Null and Alternative Hypothesis

4.1)

H_0 : The rate of female student hadn't increased. $\pi \leq 7\%$

H_1 : The rate of female student increased. $\pi > 7\%$

4.2)

H_0 : The new technique is not significantly better or worse. $\pi = 75\%$

H_1 : The new technique is significantly better or worse. $\pi \neq 75\%$

4.3)

H_0 : The machine is not dispensing too much in medium drinks. $\mu \leq 530\text{ml}$

H_1 : The machine is dispensing too much in medium drinks. $\mu > 530\text{ml}$

4.4)

H_0 : The proportion of people who received the flu shot had not changed. $\pi = 48\%$

H_1 : The proportion of people who received the flu shot had changed. $\pi \neq 48\%$

5 The p-value

F

F

T

F

F

6 Blue M&M's candies

6.1)

D

6.2)

H_0: The plain M&M's candies do contain 24% blue ones. $\pi = 24\%$

H_1: The plain M&M's candies does not contain 24% blue ones. $\pi \neq 24\%$

6.3)

The test statistic, which should be a z-statistic, because we are testing on a proportion, and $np > 10, nq > 10$.

$$z = \frac{p - \pi}{\sqrt{\frac{(1-\pi)\pi}{n}}} = \frac{18.4\% - 24\%}{\sqrt{\frac{(1-24\%)\times 24\%}{500}}} \approx -2.932$$

6.4)

The population follows a binomial distribution. Our sample satisfies $np > 10$ and $n(1-p) > 10$. Thus the test statistic approximately follows a standard normal distribution.

6.5)

There might be multiple reasons. I list some of my explanations below:

- Li Lei is very lucky and got weird M&M's candies.
- Li Lei bought the 500 candies from the same production batch, and they are not of good quality.
- The M&M might change their policy on the proportion of blue candies.
- Li Lei might have counting mistakes (e.g. He is colour blind)

6.6)

The margin of error is:

$$z_{0.025}\sqrt{\frac{(1-p)p}{n}} = 1.96 \times \sqrt{\frac{(1-0.184)\times 0.184}{500}} = 0.03$$

So the two sided 95% confidence interval is:

$$\text{C.I.} = [15.4\%, 21.4\%]$$

6.7

```
pnorm(-2.932) * 2
```

```
## [1] 0.003367867
```

We can see that the p-value is $0.003 < 0.01$. Thus, Li Lei should reject the null hypothesis.

7 Renal Disease

7.1)

C

7.2)

H_0: The mean serum-creatinine level in the people who have taken the antibiotic is not significantly different from the normal people. $\mu = 1.0\text{mg/dL}$

H_1: The mean serum-creatinine level in the people who have taken the antibiotic is significantly different from the normal people. $\mu \neq 1.0\text{mg/dL}$

7.3)

Since we are comparing between the sample and the whole normal population, and the antibiotic does not change the dispersion of the serum-creatinine level, we can assume that a z statistic should be calculated.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1.2 - 1.0}{0.4/\sqrt{12}} = \sqrt{3} = 1.732$$

7.4)

If H_0 was true, as mentioned above, the test statistics will follow a standard normal distribution.

7.5)

The margin of error is:

$$\varepsilon = z_{0.025} \times \frac{\sigma}{\sqrt{n}} = 1.96 \times \frac{0.4}{\sqrt{12}} \approx 0.2263$$

So the confidence interval is C.I. $= [0.9736, 1.4263]$.

7.6)

```
pnorm(1.732, lower.tail = F) * 2
```

```
## [1] 0.08327356
```

The \$p\$-value is 0.08 which is not smaller than 0.05, I think I will not reject H_0.

8 Serum Cholestrol Level

8.1)

From lecture 26, we now that the standard deviation in the sample is 46mg/100ml. We also know that we want a significance level of $\alpha = 0.05$. We can calculate the rejection threshold:

$$x_{\text{reject}} = \bar{x} + z_{0.05} \times \frac{\sigma}{\sqrt{n}} = 180 + 1.645 \times \frac{46}{\sqrt{50}} \approx 190.7$$

So the rejection threshold is 190.7mg/100ml.

The power, $1-\beta$, can be calculated from β, which is the possibility of not rejecting the null hypothesis when the null hypothesis is not true. First we calculate the test statistics for the population:

$$z = \frac{x_{\text{reject}} - \mu}{\sigma/\sqrt{n}} = \frac{190.7 - 200}{46/\sqrt{50}} \approx -1.43$$

Then we can calculate the probability of $P(\text{not reject } H_0 | H_0 \text{ is false})$ which is β.

```
pnorm(-1.43)
```

```
## [1] 0.07635851
```

So $\beta = 0.076$. The power is $1 - \beta = 0.924$.

8.2)

Because I don't want a step by step calculation, so I'm just going to derive the general formula.

The test statistic we calculate are z_α and $z_{1-\beta} = -z_\beta$. From the previous example, we can see that:

$$-z_\beta = \frac{\bar{x} + z_\alpha \times \frac{\sigma}{\sqrt{n}} - \mu}{\sigma/\sqrt{n}}$$

With a bit of fumbling around, we can get the equation:

$$n = (\frac{(z_\alpha + z_\beta)\sigma}{\mu - \bar{x}})^2$$

So in our case, $\lceil n \rceil = 101$, so the required sample size is 101.

8.3)

B

9 Type I & II errors

9.1)

B

9.2)

C

9.3)

B

9.4)

D (But I think C is also ok, because we don't want to waste a lot of money to clean the pool.)