

# STAT-UB 103 Homework 7

Ishan Pranav

March 30, 2023

## 1 A random sample

$$\varphi(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}.$$

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

$$P(z \leq x) = \Phi(x).$$

$$\text{probit}(p) = \Phi^{-1}(p) \text{ for } p \in (0, 1).$$

Let  $n$  represent the number of samples,  $\sigma_X$  represent the standard deviation,  $\bar{X}$  represent the sample mean,  $\alpha$  represent the significance level, and  $H_0$  represent the null hypothesis.  $X$  can be modeled with a Gaussian distribution.

$$n = 100.$$

$$\sigma_X = 60.$$

$$\bar{X} = 110.$$

$$\alpha = 0.05.$$

$$H_0 : \mu_X = 100.$$

a. Let  $H_1$  represent the alternative hypothesis.

$$H_1 : \mu_X > 100.$$

$$\begin{aligned}
P(\bar{X} \geq 110 \mid H_0) &= P\left(z \geq \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \mid \mu_{\bar{X}} = 110\right) \\
&= P\left(z \geq \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \mid \mu_X = 110\right) \\
&= P\left(z \geq \frac{110 - 100}{\frac{60}{\sqrt{100}}}\right) \\
&= 1 - P\left(z \leq \frac{10}{6}\right).
\end{aligned}$$

$$1 - \Phi\left(\frac{10}{6}\right) \approx 0.0478 \dots$$

Given that the true mean of the population is 100, the probability of a sample mean greater than or equal to 110 is less than 0.05. There is sufficient evidence, at the 5-percent significance level, to conclude that the true mean of the population is greater than 100.

b. Let  $H_1$  represent the alternative hypothesis.

$$H_1 : \mu_X \neq 100.$$

$$\begin{aligned}
z^* &= \text{probit}(\alpha) \\
&= \text{probit}\left(\frac{0.05}{2}\right) \\
&= \text{probit}(0.025) \\
&\approx 1.960 \dots
\end{aligned}$$

$$-1.960 \approx \frac{\bar{X}_{\min} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \approx \frac{\bar{X}_{\min} - 100}{\frac{60}{\sqrt{100}}}.$$

$$\bar{X}_{\min} \approx 88.2402 \dots$$

$$1.960 \approx \frac{\bar{X}_{\max} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \approx \frac{\bar{X}_{\max} - 100}{\frac{60}{\sqrt{100}}}.$$

$$\bar{X}_{\max} \approx 111.760 \dots$$

The 95% confidence interval for  $\mu_X$  is (88.2402, 111.760).

We are 95 percent confident that the true mean of the population is between approximately 88.24 and approximately 111.76. There is insufficient evidence, at the 5-percent significance level, to conclude that the true mean of the population is different from 100.

- c. The results differ because the significance level is stricter when applied to a two-tailed test. This is because the probability is divided by two and distributed to both extremes of the distribution. This means that only sample means in one of the two 2.5-percent-area regions are statistically significant. Meanwhile, the one-tailed test is more liberal: It treats any sample mean in the 5-percent-area region as statistically significant.

## 2 Trading skills of institutional investors

$$f_{\nu}(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}.$$

$$F_{\nu}(t) = \int_{-\infty}^t f_{\nu}(u) du.$$

Let  $X$  represent the rate of return,  $\bar{X}$  represent the sample mean,  $\mu_X$  represent the true mean,  $s_X$  represent the standard error, and  $\nu$  represent the degrees of freedom.  $X$  can be modeled with a Student distribution.

$$\bar{X} = 2.95\%.$$

$$n = 200.$$

$$\nu = n - 1 = 199.$$

- a. Let  $H_0$  represent the null hypothesis and  $H_1$  represent the alternative hypothesis.

$$H_0 : \mu_X = 0.$$

$$H_1 : \mu_X > 0.$$

- b. Let  $\alpha$  represent the significance level.

$$\alpha = 0.05.$$

$$t^* = F_{\nu}^{-1}(\alpha) = F_{199}^{-1}(0.05) = 1.6525 \dots$$

$$1.6525 \approx \frac{\bar{X}_{\max} - \mu_{\bar{X}}}{s_{\bar{X}}} \approx \frac{\bar{X}_{\max}}{\frac{8.82\%}{\sqrt{200}}}.$$

$$\bar{X}_{\max} \approx 1.0306 \dots \%$$

The rejection region is  $(1.0306, \infty)$ .

- c. Assume that the hypothesis is true: The true mean rate of return of round-trip trades is not positive. If a specific positive sample mean with a probability less than 5% actually occurs, then that sample is statistically significant at the  $\alpha = 0.05$  level, and, by contradiction, it provides convincing evidence against the hypothesis. In other words,  $\alpha$  is the probability of making a false positive (Type I) error: showing that institutional investors performed successfully when in fact they did not.
- d. In the given test output, the test statistic ( $t$ ) is 4.73, and the probability ( $P$ -value) is 0.000.
- e. Given that the true mean rate of return of round-trip trades is not positive, the probability of a sample mean greater than or equal to 2.95 percent is effectively zero. There is sufficient evidence, at the 5-percent significance level, to conclude that the true mean rate of return is positive.

### 3 A new dental bonding agent

- a. Let  $H_0$  represent the null hypothesis and  $H_1$  represent the alternative hypothesis.

$$H_0 : \mu_X = 5.70 \text{ MPa.}$$

$$H_1 : \mu_X < 5.70 \text{ MPa.}$$

- b. Let  $\alpha$  represent the significance level,  $n$  represent the number of samples,  $\nu$  represent the degrees of freedom, and  $s_X$  represent the standard error.

$$\alpha = 0.01.$$

$$n = 10.$$

$$\nu = n - 1 = 9.$$

$$s_X = 0.46 \text{ MPa.}$$

$$t^* = F_\nu^{-1}(\alpha) = F_9^{-1}(0.01) \approx -2.8203 \dots$$

$$-2.8203 \approx \frac{\bar{X}_{\min} - \mu_{\bar{X}}}{s_{\bar{X}}} \approx \frac{\bar{X}_{\min} - 5.70 \text{ MPa}}{\frac{0.46 \text{ MPa}}{\sqrt{10}}}.$$

$$\bar{X}_{\min} \approx 5.29 \text{ MPa.}$$

The rejection region is  $(-\infty, 5.29 \text{ MPa})$ .

- c. Let  $\bar{X}$  represent the sample mean and  $t$  represent the test statistic.

$$\bar{X} = 5.07 \text{ MPa.}$$

$$t = \frac{5.07 \text{ MPa} - 5.70 \text{ MPa}}{\frac{0.46 \text{ MPa}}{\sqrt{10}}} \approx -4.3309.$$

$$P = 0.0010 \dots$$

- d. Given that the true mean is 5.70 MPa, the probability of a sample mean less than or equal to 5.07 MPa is approximately 0.0010. There is sufficient evidence, at the 5-percent significance level, to conclude that the true mean is less than 5.70 MPa.
- e. The test results are only valid if the samples are random, independent, and approximately normally distributed with approximately homogeneous variances.

## 4 Satellite radio in cars

- a. The parameter of interest is  $\pi$ , the true proportion of all satellite radio subscribers who have a satellite radio receiver in their car.
- b. Let  $H_0$  represent the null hypothesis.

$$H_0 : \pi = 80\%.$$

- c. Let  $H_1$  represent the alternative hypothesis.

$$H_1 : \pi < 80\%.$$

- d. Let  $\alpha$  represent the significance level,  $n$  represent the sample size,  $\nu$  represent the degrees of freedom,  $p$  represent the sample proportion,  $s_p$  represent the standard error, and  $t$  represent the test statistic.

$$\alpha = 0.10.$$

$$n = 501.$$

$$\nu = n - 1 = 500.$$

$$p = \frac{396}{501}.$$

$$s_p = \sqrt{\frac{(p)(1-p)}{n}}.$$

$$t = \frac{\frac{396}{501} - 80\%}{\sqrt{\frac{(80\%)(20\%)}{501}}} \approx -0.54.$$

$$t^* = F_\nu^{-1}(\alpha) = F_{500}^{-1}(0.10) \approx -1.3.$$

e.

$$-1.3 \approx \frac{p_{\min} - \mu_p}{s_p} \approx \frac{p_{\min} - 80\%}{\sqrt{\frac{(80\%)(20\%)}{396}}}.$$

$$p_{\min} \approx 77\%.$$

The rejection region is  $(-\infty, 77\%)$ .

f. Let  $P$  represent the probability of the given test statistic.

$$P \approx 0.2961 \dots$$

g. Given that the true proportion is 80 percent, the probability of a sample proportion less than or equal to  $\frac{396}{501}$  is approximately 0.2961. There is insufficient evidence, at the 10-percent significance level, to conclude that the true mean is less than 80 percent. The rejection-region method corroborates this result:  $\frac{396}{501} \approx 79\%$  is greater than the minimum sample proportion (77%) and thus outside the rejection region.

## 5 Test

Let  $H_0$  represent the null hypothesis,  $H_1$  represent the alternative hypothesis,  $\alpha$  represent the significance level,  $\mu_X$  represent the true mean and the parameter of interest,  $\sigma_X$  represent the standard deviation,  $n$  represent the sample size, and  $\bar{X}$  represent the sample mean.

$$H_0 : \mu_X = 500.$$

$$H_1 : \mu_X > 500.$$

$$\alpha = 0.05.$$

$$\sigma_X = 100.$$

$$n = 25.$$

a. Assuming that the null hypothesis is true, the sampling distribution of the sample mean has a mean  $\mu_{\bar{X}} = 500$  and a standard deviation  $\sigma_{\bar{X}} = \frac{100}{\sqrt{25}} = 20$ .

b. Let  $\bar{X}_0$  represent the value of  $\bar{X}$  above which the null hypothesis would be rejected.

$$z^* = \text{probit}(\alpha) = \text{probit}(0.05) \approx 1.6449 \dots$$

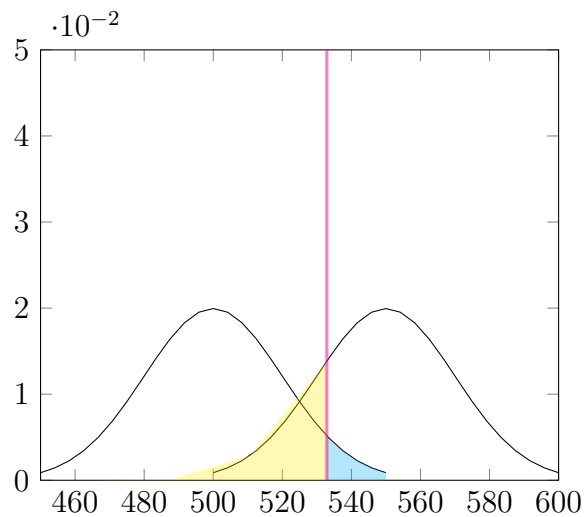


Figure 1: The sampling distribution of the sample mean. The cyan-shaded region represents  $\alpha$ . The yellow-shaded region represents  $\beta$ . The magenta axis represents the decision boundary.

$$1.6449 \approx \frac{\bar{X}_0 - \mu_{\bar{X}}}{s_{\bar{X}}} \approx \frac{\bar{X}_0 - 500}{20}.$$

$$\bar{X}_0 \approx 532.897 \dots$$

c. See Figure 1.

d. Let  $\beta$  represent the probability of a false negative (Type II error).

$$\begin{aligned} \beta &= P(\bar{X} < \bar{X}_0 \mid \mu_{\bar{X}} = 550) \\ &= P\left(z < \frac{\bar{X}_0 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(z < \frac{532.897 - 550}{20}\right) \\ &= P(z < -0.8551 \dots) \\ &\approx 0.1962 \dots \end{aligned}$$

e.

$$1 - \beta \approx 0.8037 \dots$$

## 6 Refer

- a. The probability that the hypothesis test will incorrectly fail to reject the null hypothesis is 1.76 percent.

$$\begin{aligned}\beta &= P(\bar{X} < \bar{X}_0 \mid \mu_{\bar{X}} = 575) \\ &= P\left(z < \frac{\bar{X}_0 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(z < \frac{532.897 - 575}{20}\right) \\ &= P(z < -2.1051 \dots) \\ &\approx 0.0176 \dots\end{aligned}$$

- b. The power of the test is 98.2 percent.

$$1 - \beta \approx 0.9824 \dots$$

- c. The power of the test is higher when  $\mu_{\bar{X}} = 575$  than when  $\mu_{\bar{X}} = 550$ . The three ways to increase power are to increase the significance level ( $\alpha$ ), to increase the mean difference, or to decrease the standard deviation. In this case, the increase in the mean difference accounts for the increase in statistical power.