Suggested problems Ch 7

The mean serum-creatinine level measured in 12 patients 24 hours after they received a newly proposed antibiotic was 1.2 mg/dL.

***7.1** If the mean and standard deviation of serum creatinine in the general population are 1.0 and 0.4 mg/dL, respectively, then, using a significance level of .05, test whether the mean serum-creatinine level in this group is different from that of the general population.

***7.2** What is the p-value for the test?

***7.3** Suppose the sample standard deviation of serum creatinine in Problem 7.1 is 0.6 mg/dL. Assume that the standard deviation of serum creatinine is not known, and perform the hypothesis test in Problem 7.1. Report a p-value.

***7.4** Compute a two-sided 95% CI for the true mean serum-creatinine level in Problem 7.3.

***7.5** How does your answer to Problem 7.4 relate to your answer to Problem 7.3.

7.6 Suppose $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} = -1.52$ and a one-sample t test is performed based on seven subjects. What is the two-tailed p-value?

7.7 Use a computer program to compute the probability that a t distribution with 36 df exceeds 2.5.

7.8 Use a computer program to compute the lower 10th percentile of a t distribution with 54 df.

7.1 We test the hypothesis $H_0: \mu = 1.0$ versus $H_1: \mu \neq 1.0$, where $\sigma = 0.4$ under either hypothesis. We use the one-sample z-test. The rejection region is defined by $z < z_{.025} = -1.96$ or $z > z_{.975} = 1.96$, where

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{1.2-1.0}{0.4/\sqrt{12}} = \frac{0.2}{0.1155} = 1.732$$

Since $-1.96 < 1.732 < 1.96$, it follows that we accept H_0 at the 5% level.

7.2 Since this is a two-sided test and $z > 0$, the p-value is given by

$$p = 2 \times [1 - \Phi(1.732)] = 2 \times (1 - .9584) = .083$$

7.3 We use a one-sample t test. The rejection region is defined by $t < t_{11,.025}$ or $t > t_{11,.975}$. We have the test statistic

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{1.2-1.0}{0.6/\sqrt{12}} = \frac{0.2}{0.173} = 1.155$$

Since $t_{11,.025} - 2.201$, $t_{11,.975} = 2.201$ and $-2.201 < 1.155 < 2.201$, it follows that we accept H_0 at the 5% level. Furthermore, to obtain the p-value we compute $2 \times Pr(t_{11} > t)$.

Since $t_{11,.85} = 1.088$, $t_{11,.9} = 1.363$ and $1.088 < 1.155 < 1.363$, it follows that $1-.9 < \frac{p}{2} < 1-.85$ or $.1 < \frac{p}{2} < .15$ or $.2 < p < .3$. The exact p-value $= .27$.

7.4 A two-sided 95% confidence interval is given by

$$\begin{aligned}\bar{x} \pm t_{n-1,\ .975}\frac{s}{\sqrt{n}} &= 1.2 \pm t_{11,\ .975}\frac{0.6}{\sqrt{12}} \\ &= 1.2 \pm \frac{2.201(0.6)}{\sqrt{12}} \\ &= 1.2 \pm 0.38 = (0.82,\ 1.58)\end{aligned}$$

7.5 This interval contains 1.0 = the mean for the general population. This is consistent with the result in Problem 7.3 where we accepted H_0 using a two-sided test at the 5% level.

7.6 $p = 2 \times Pr(t_6 < -1.52) = 2 \times Pr(t_6 > 1.52)$. Since $t_{6,.9} = 1.440$, $t_{6,.95} = 1.943$, and $1.440 < 1.52 < 1.943$, it follows that $2 \times (1-.95) < p < 2 \times (1-.9)$, or $.1 < p < .2$. The exact p-value, obtained by computer, is $p = .179$.

7.7 The probability that a t distribution with 36 df exceeds 2.5 is $\Pr(t_{36,.05} > 2.5) = 1 - P(t_{36,.05} \le 2.5)$. In R, we have:

```
> 1-pt(2.5, 36)
[1] 0.008556915
```

7.8 The lower 10th percentile of a t distribution with 54 df is

```
> 1-qt(.10, 54)
[1] -1.297426
```

Suppose the incidence rate of myocardial infarction (MI) was 5 per 1000 among 45- to 54-year-old men in 2000. To look at changes in incidence over time, 5000 men in this age group were followed for 1 year starting in 2010. Fifteen new cases of MI were found.

7.12 Using the critical-value method with $\alpha = .05$, test the hypothesis that incidence rates of MI changed from 2000 to 2010.

7.13 Report a p-value to correspond to your answer to Problem 7.12.

Suppose that 25% of patients with MI in 2000 died within 24 hours. This proportion is called the 24-hour case-fatality rate.

7.14 Of the 15 new MI cases in the preceding study, 5 died within 24 hours. Test whether the 24-hour case-fatality rate changed from 2000 to 2010.

7.15 Suppose we eventually plan to accumulate 50 MI cases during the period 2010–2015. Assume that the 24-hour case-fatality rate is truly 20% during this period. How much power would such a study have in distinguishing between case-fatality rates in 2000 and 2010–2015 if a two-sided test with significance level .05 is planned?

7.16 How large a sample is needed in Problem 7.15 to achieve 90% power?

7.12 We wish to test the hypothesis $H_0: p = p_0$ versus $H_1: p \neq p_0$, where $p_0 = .005$, $p =$ true incidence rate of MI in 2010 among 45–54-year-old men. Since $np_0q_0 = 5000(.005)(.995) = 24.88 \geq 5$, we can use the normal-theory method. The rejection region is given by $z > z_{1-\alpha/2}$ or $z < z_{\alpha/2}$, where

$$\begin{aligned} z &= \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \\ &= \frac{15/5000 - .005}{\sqrt{.005(.995)/5000}} = \frac{-.002}{.000997} \\ &= -2.005 \end{aligned}$$

Since $z = -2.005 < z_{\alpha/2} = z_{.025} = -1.96$, we reject H_0 at the 5% level.

7.13 The p-value is given by $2\Phi(z) = 2\Phi(-2.005) = 2 \times [1 - \Phi(2.005)] = 2(1 - .9775) = .045$.

7.14 We wish to test the hypothesis $H_0: p = p_0$ versus $H_1: p \neq p_0$, where $p_0 = .25$. Since $np_0q_0 = 15(.25)(.75) = 2.81 < 5$, we must use the exact method to test these hypotheses. Since $\hat{p} = \frac{5}{15} = .333 > p_0$, the two-tailed p-value is obtained from

$$\begin{aligned} p &= 2 \times \sum_{k=5}^{15} {}_{15}C_k (.25)^k (.75)^{15-k} \\ &= 2 \times \left(1 - \sum_{k=0}^{4} {}_{15}C_k (.25)^k (.75)^{15-k}\right) \end{aligned}$$

We refer to the exact binomial tables (Table 1) under $n = 15$, $p = .25$, and obtain $Pr(0) = .0134$, $Pr(1) = .0668$, $Pr(2) = .1559$, $Pr(3) = .2252$, $Pr(4) = .2252$. Thus,

$$\begin{aligned} p &= 2 \times [1 - (.0134 + \ldots + .2252)] \\ &= 2 \times (1 - .6865) = .627 \end{aligned}$$

Therefore, there is no significant change in the case-fatality rate between 2000 and 2010.

7.15 We use the power formula in Equation 7.32 (in Chapter 7, text) using a two-sided formulation whereby

$$\text{Power} = \Phi\left[\sqrt{\frac{p_0 q_0}{p_1 q_1}}\left(z_{\alpha/2} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_0 q_0}}\right)\right]$$

where $p_0 = .25$, $p_1 = .20$, $\alpha = .05$, $n = 50$. We have

$$\text{Power} = \Phi\left[\sqrt{\frac{.25(.75)}{.20(.80)}}\left(z_{.025} + \frac{|0.25 - 0.20|\sqrt{50}}{\sqrt{.25(.75)}}\right)\right]$$

$$= \Phi\left[\sqrt{\frac{.1875}{.16}}\left(-1.96 + \frac{.05\sqrt{50}}{\sqrt{.1875}}\right)\right]$$

$$= \Phi[1.0825(-1.96 + 0.8165)] = \Phi(-1.238)$$

$$= 1 - \Phi(1.238) = 1 - .89 = .11$$

Thus, such a study would only have an 11% chance of detecting a significant difference.

7.16 To compute the sample size needed to achieve 90% power, we use the formula in Equation 7.33 (in Chapter 7, text) using a two-sided formulation whereby

$$n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta}\sqrt{\frac{p_1 q_1}{p_0 q_0}}\right)^2}{(p_1 - p_0)^2}$$

$$= \frac{.25(.75)\left[z_{.975} + z_{.90}\sqrt{\frac{.20(.80)}{.25(.75)}}\right]^2}{(.20 - .25)^2}$$

$$= \frac{.1875[1.96 + 1.28(0.9238)]^2}{.0025}$$

$$= 75(3.1424)^2 = 740.6$$

Thus, we would need to study 741 MI cases to achieve 90% power.

Ribosomal 5S RNA can be represented as a sequence of 120 nucleotides. Each nucleotide can be represented by one of four characters: A (adenine), G (guanine), C (cytosine), or U (uracil). The characters occur with different probabilities for each position. We wish to test whether a new sequence is the same as ribosomal 5S RNA. For this purpose, we replicate the new sequence 100 times and find there are 60 A's in the 20th position.

7.21 If the probability of an A in the 20th position in ribosomal 5S RNA is .79, then test the hypothesis that the new sequence is the same as ribosomal 5S RNA using the critical-value method.

7.22 Report a *p*-value corresponding to your results in Problem 7.21.

Suppose we wish to test the hypothesis H_0: $\mu = 45$ vs. H_1: $\mu > 45$.

7.23 What will be the result if we conclude that the mean is greater than 45 when the actual mean is 45?

(i) We have made a type I error.

(ii) We have made a type II error.

(iii) We have made the correct decision.

7.24 What will be the result if we conclude that the mean is 45 when the actual mean is 50?

(i) We have made a type I error.

(ii) We have made a type II error.

(iii) We have made the correct decision.

Suppose we wish to test H_0: $\mu = 30$ vs. H_1: $\mu \neq 30$ based on a sample of size 31.

7.25 Which of the following sample results yields the smallest *p*-value and why?

(i) $\bar{x} = 28, s = 6$

(ii) $\bar{x} = 27, s = 4$

(iii) $\bar{x} = 32, s = 2$

(iv) $\bar{x} = 26, s = 9$

7.21 We wish to test the hypothesis $H_0: p = p_0 = .79$ versus $H_1: p \neq p_0$. Since $np_0q_0 = 100(.79)(.21) = 16.6 \geq 5$, we can use the normal-theory method. The critical values are given by $z_{.975} = 1.96$, where we use a 5% level of significance. We reject H_0 if $z_{corr} > 1.96$, where z is given by

$$z_{corr} = \frac{|.60-.79| - \frac{1}{2\times 100}}{\sqrt{.79(.21)/100}}$$

$$= \frac{.19-.005}{.0407} = 4.545$$

Since $z_{corr} = 4.545 > 1.96$, we reject H_0 at the 5% level.

7.22 The two-tailed p-value $= 2\times\left[1-\Phi(z_{corr})\right] = 2\times\left[1-\Phi(4.545)\right] = 5.49\times 10^{-6}$.

7.23 If we conclude that the mean is greater than 45 when it actually is equal to 45, then we have made a type I error.

7.24 If we conclude that the mean is 45 when the actual mean is 50, then we have made a type II error.

7.25 We have the following test statistics for (i) – (iv):

(i) $t = \frac{28\text{-}30}{6/\sqrt{31}} = \frac{-2}{1.078} = -1.86$

(ii) $t = \frac{27\text{-}30}{4/\sqrt{31}} = \frac{-3}{0.718} = -4.18$

(iii) $t = \frac{32\text{-}30}{2/\sqrt{31}} = \frac{2}{0.359} = 5.57$

(iv) $t = \frac{26\text{-}30}{9/\sqrt{31}} = \frac{-4}{1.616} = -2.47.$

The test statistic which is largest in absolute value will have the smallest p-value. This is the test statistic in (iii) (i.e., 5.57).