

Applied Statistics - MATH 7343

Homework-4

Sai Nikhil

NUID: 001564864

Sai Nikhil



①

For A, $p\text{-value} = 0.01 < \alpha = 0.05$

For B, $p\text{-value} = 0.009 < \alpha = 0.05$

→ We reject null hypothesis for both A&B.

Also, $A_{H_0A} - p\text{-value} > H_{0B} - p\text{-value}$

→ Statistic for H_{0A} is close to mean that for H_{0B}

→ Mutation B has greater chance of
true cause of high cholesterol level.

②

10.8.2

No, there isn't a way to prove a null hypothesis. We can only report it with a certain level of confidence. For any fixed, finite sample size, there will always be some small but nonzero effect size for which the statistical test has virtually no power. More practically, though, one can prove that within some epsilon of null hypothesis, the deviations less than this epsilon are not practically significant.

10.8.4

There is a close relationship between Confidence interval and hypothesis testing. When constructing a 95% CI, all values in the interval are considered plausible for the parameter being estimated. The values outside this interval are rejected as relatively implausible. If the value of a parameter specified by the null hypothesis is contained in the 95% interval, then the null hypothesis cannot be rejected at the levels of 0.05 levels. If the value specified by the null hypothesis is not in the interval, then the null hypothesis can be rejected at the 0.05 level. If a 99% confidence interval is constructed, then values outside the interval are rejected at the 0.01 level.

10.8.13

Technically speaking, the probability of occurrence of a type II error is the probability of failing to reject the null hypothesis when it is false. The only way to make this probability equal to 0 is to always reject every null hypothesis, which is not a practical solution. Hence, it is not possible for FDA to completely eliminate the occurrence of type II errors.

③

10-8-10

(a) For a one-sided test,
Null Hypothesis (H_0)

Mean WBC count (μ) = $7250 / \text{mm}^3$

Alternate Hypothesis (H_A)

Mean WBC count (μ) $< 7250 / \text{mm}^3$

(b) $\bar{x} = 4767 / \text{mm}^3$, $\sigma = 3204 / \text{mm}^3$

$$\alpha = 0.05$$

$$t_0 = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{4767 - 7250}{\frac{3204}{\sqrt{15}}}$$

$$\Rightarrow t_0 \approx -3.0014$$

$$t_{0.05, 15-4} = t_{0.05, 14} = -1.761 > t_0$$

\Rightarrow we reject H_0

(c)

The hypotheses test concludes that,
there is enough evidence to support
the alternate hypothesis i.e.,

mean WBC count (μ) $< 7250/\text{mm}^3$

(4)

10.8.14

(a)

$$P(\text{Type I error}) = \alpha = 0.05$$

(b)

Hypotheses: $H_0: \mu \geq 244$ $H_A: \mu < 244$

$$P(\text{Type II error}) = P(\text{Accept } H_0 \mid H_A \text{ is true})$$

$$\alpha = 0.05, \quad Z_\alpha = 1.645$$

$$\bar{x} = \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}} = 244 - 1.645 \times \frac{41}{\sqrt{25}}$$

$$\approx 230.51$$

$$\text{At } \mu_1 = 219, \quad z = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{230.51 - 219}{\frac{41}{\sqrt{25}}} \approx 1.40$$

$$\approx 1.40$$

$$\therefore P(\text{Type II error})(\beta) = P(z > 1.40) = 0.081$$

c) Power = $1 - \beta = 0.919$

d) The power can be increased by increasing the sample size.

e) $P(\text{Type II error}) = P(Z > z_{-1.64}) \leq 0.05$

$$\Rightarrow -1.64 + \left(\frac{244 - 219}{41} \right) \sqrt{n} > 1.64$$

$$n > 28.94$$

$$n \geq 29$$

f)

$$P\left(Z > -1.64 + \left(\frac{244 - 219}{41} \right) \sqrt{n}\right) \leq 0.10$$

$$-1.64 + \left(\frac{244 - 219}{41} \right) \sqrt{n} > 1.28$$

$$n > 22.93$$

$$n \geq 23$$

5

10.8.16

Code:

```
> library(psych)
>
> heart.data <- read.table(file="heart.txt", header=TRUE, na.strings=".") 
>
> describeBy(heart.data$pdi, heart.data$trtment)

  Descriptive statistics by group
group: 0
  vars n mean sd median trimmed mad min max range skew kurtosis se
X1   1 73 91.92 16.49      92  91.97 17.79  60 134     74  0.09 -0.39 1.93
-----
group: 1
  vars n mean sd median trimmed mad min max range skew kurtosis se
X1   1 70 97.77 14.69      98  98.41 9.64   50 124     74 -0.65  1.25 1.76
> describeBy(heart.data$pdi, heart.data$trtment)

  Descriptive statistics by group
group: 0
  vars n mean sd median trimmed mad min max range skew kurtosis se
X1   1 73 91.92 16.49      92  91.97 17.79  60 134     74  0.09 -0.39 1.93
-----
group: 1
  vars n mean sd median trimmed mad min max range skew kurtosis se
X1   1 70 97.77 14.69      98  98.41 9.64   50 124     74 -0.65  1.25 1.76
>
> describe(heart.data$pdi)
  vars n mean sd median trimmed mad min max range skew kurtosis se
X1   1 143 94.78 15.85      98  95.27 16.31  50 134     84 -0.26  0.06 1.33
> describe(heart.data$mdi)
  vars n mean sd median trimmed mad min max range skew kurtosis se
X1   1 144 104.74 15.6    106.5 105.66 14.08  50 142     92 -0.67  1.06 1.3
>
> x <- heart.data$pdi[!is.na(heart.data$pdi)]
> y <- heart.data$mdi[!is.na(heart.data$mdi)]
>
> t.test(x, alternative = "two.sided", mu = 100)

  One Sample t-test

data: x
t = -3.9356, df = 142, p-value = 0.0001294
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
 92.16289 97.40354
sample estimates:
mean of x
 94.78322

> t.test(y, alternative = "two.sided", mu = 100)

  One Sample t-test

data: y
t = 3.6421, df = 143, p-value = 0.0003773
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
 102.1657 107.3065
sample estimates:
mean of x
 104.7361
```

[a], [c]

[b], [c]

$$(a) \text{ p-value} = 0.00013 < 0.05$$

\Rightarrow Reject null Hypotheses (H_0)

\therefore Mean PDI score for children born with Congenital heart disease who undergo reparative heart surgery during first 3months of life is not equal to 100.

$$(b) \text{ p-value} = 0.00038 < 0.05$$

\Rightarrow Reject null hypotheses (H_0)

\therefore Mean MDI score for children born with Congenital heart disease who undergo reparative heart surgery during first 3months of life is not equal to 100

$$(c) 95\%-CI \text{ for PDI} \approx 95\%-CI \text{ for MDI}$$

both do not contain 100.

It is expected that they don't contain

100 because we rejected both.

6

```
> beds <- read.table(file="bed.txt", header=TRUE)
> library(psych)
>
> describe(beds$bed80)
   vars n mean   sd median trimmed  mad min max range skew kurtosis    se
X1     1 51 4.56 1.01     4.5     4.51 1.04 2.7 7.4    4.7  0.6     0.32 0.14
> describe(beds$bed86)
   vars n mean   sd median trimmed  mad min max range skew kurtosis    se
X1     1 51 4.23 1.11     4.2     4.13 1.19 2.4 7.7    5.3  0.92     0.99 0.16
>
> t.test(beds$bed80, beds$bed86, var.equal = TRUE)

  Two Sample t-test

data: beds$bed80 and beds$bed86
t = 1.5396, df = 100, p-value = 0.1268
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.09337022  0.74042905
sample estimates:
mean of x mean of y
4.556863  4.233333

> t.test(beds$bed80, beds$bed86, paired = T)

  Paired t-test

data: beds$bed80 and beds$bed86
t = 6.8721, df = 50, p-value = 9.513e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2289696 0.4180892
sample estimates:
mean of the differences
  0.3235294
>
```

(a)

Descriptive Statistics

(b)

pvalue = 0.1268 > 0.05

there is no sufficient evidence to indicate that

mean of beds per 1000 population in

the years 1980 & 1986 are different

(c) $p\text{-value} \approx 0 < 0.05$

\Rightarrow there is sufficient evidence to indicate that number of beds per 1000 population in the years 1980 & 1986 are different

(d) The difference is that in (b) each sample is taken from different population independently and in (c) each sample is taken from one single population and paired. Depending on the p-value obtained we can say that the result is different in both cases.

(e) 95% CI for Two sample t-test is,

$$(-0.0934, 0.7404)$$

95% CI for paired sample t-test is,

$$(0.2290, 0.4181)$$