Introductory Statistics Tutorial Answers

Chapter 9 – Significance Testing: Using Data to Test Hypotheses

Section A: Quiz

1. The null hypothesis is the hypothesis tested by the statistical test. The alternative hypothesis specifies the type of departure from the null hypothesis we expect to detect.

2. **(a)** $H_0 : \theta \neq \theta_0$ **(b)** $H_0 : \theta > \theta_0$ **(c)** $H_0 : \theta < \theta_0$

3. A one-tailed test is used when the investigators have good grounds for believing the true value of θ was on one particular side of θ_0 before the study began. Otherwise, or if in doubt, a two-tailed test is used. Good grounds mean that there is prior information or there is a theory to tell the investigators which way the study will go.

4. $t_0 = \dfrac{\text{estimate - hypothesised value}}{\text{std error}}$

5. The *P-value* is the probability that, if the **null hypothesis** was true, sampling variation would produce an estimate that is at least as far away from the hypothesised value as our data estimate.

6. The *P-value* measures the **strength** of evidence against the null hypothesis.

7.

P-value	**Evidence against H_0**
> 0.12	**none**
≈ 0.10	**weak**
≈ 0.05	**some**
≈ 0.01	**strong**
≤ 0.001	**very strong**

8. Nothing.

9. A confidence interval.

10. *P-value* < 5%

Section B: Doing Tests by Hand

1. **(a)** Let p_W be the true proportion of white prisoners who were infected with TB and p_G be the true proportion of Gypsy prisoners who were infected with TB. Thus $\theta = p_W - p_G$.

(b) H_0: p_W - p_G = 0 vs H_1: p_W - $p_G \neq 0$

(c) $\hat{p}_W - \hat{p}_G = \dfrac{496}{886} - \dfrac{74}{152} = 0.5598 - 0.4868 = 0.0730$

(d) $\text{se}(\hat{p}_W - \hat{p}_G) = \sqrt{\dfrac{0.5598(1-0.5598)}{886} + \dfrac{0.4868(1-0.4868)}{152}} = 0.04384$

$z_0 = \dfrac{0.0730 - 0}{0.04384} = 1.665$

(e) *P-value* = 2 x pr($Z > 1.665$)

= between 0.05 and 0.1 (in fact it is just less than 0.1)

(f) We have weak evidence against H_0.

(g) There is weak evidence that there is a difference between the proportion of White prisoners who had TB and the proportion of Gypsy prisoners who had TB.

(h) 95% confidence interval for $p_W - p_G$:

0.0730 ± 1.96 x $0.04384 = (-0.013, 0.159)$

(i) With 95% confidence, we estimate that the proportion of White prisoners who had TB is somewhere between 0.013 lower than and 0.159 higher than the proportion of Gypsy prisoners who had TB.

Section C: Interpreting Output and Interpretation Issues

1. **(a)** Let μ_1 be the true mean daily revenue for laundry 1 and μ_2 be the true mean daily revenue for laundry 2. Thus the parameter used is $\mu_1 - \mu_2$, the difference in the mean daily revenue for the two laundries.

(b) H_0: μ_1 - μ_2 = 0 vs H_1: μ_1 - $\mu_2 \neq 0$

(c) $\bar{x}_1 - \bar{x}_2 = 635.4 - 601.6 = 33.8$

(d) $t_0 = 1.94$

(e) *P-value* = 0.057. We have some evidence that the mean daily revenue of the first laundry is greater than the mean daily revenue of the second laundry.

(f) With 95% confidence, we estimate that the mean daily revenue of the first laundry is somewhere between \$1 less than and \$69 more than the mean daily revenue of the second laundry.

(g) The computer uses a different formula for calculating *df*. This formula gives a larger value of *df* than the hand calculation based on the minimum of one less than each sample size.

2. (1)

3. (1)

4. (4)