MS Module 6: Hypothesis testing of proportions – practice problems

(The attached PDF file has better formatting.)

Exercise 6.1: Proportions, probability of a Type II error, and required observations

The average proportion of death from a disease is 80%. A study tests whether a drug reduces the proportion of death from the disease. Of 500 subjects who are given the drug, 380 die from the disease. Let μ be the expected proportion of death from the disease among subjects given the drug.

- The null hypothesis is H_0: the expected proportion of subjects dying $\mu = \mu_0 = 80\%$
- The two-tailed alternative hypothesis is H_a: the expected proportion of subjects dying $\mu \neq \mu_0$ (80%)
- The one-tailed alternative hypothesis is H_a: the expected proportion dying $\mu < \mu_0$ (80%)

A. What is the incidence of death from the disease in the sample of 500 subjects?
B. What is the standard deviation of the incidence of death in the sample if the null hypothesis is true?
C. What is the *z* value used to test the null hypothesis?
D. What is the *p* value for the one-tailed alternative hypothesis?
E. What is the *p* value for the two-tailed alternative hypothesis?
F. What are the expected value and variance of the *z* value for testing the null hypothesis?

If the null hypothesis is tested at a 2% significance level and the true incidence of death with the drug is 75%:

G. What is the expected value of the *z* value for testing the null hypothesis?
H. What is the variance of the *z* value for testing the null hypothesis?
I. What is the probability of a Type II error for the one-tailed test?
J. What is the probability of a Type II error for the two-tailed test?
K. How many observations are needed for a one-tailed test if $\alpha = 2\%$ and $\beta = 10\%$?
L. How many observations are needed for a two-tailed test if $\alpha = 2\%$ and $\beta = 10\%$?

Part A: The incidence of death from the disease in the sample of 500 subjects is 380 / 500 = 76%.

Part B: The standard deviation of the incidence of death in the sample if the null hypothesis is true =

$(p\,(1 - p) / N)^{½} = (80\% \times (1 - 80\%) / 500\,)^{0.5} = 0.0178885$.

Part C: The *z* value used to test the null hypothesis = (76% – 80%) / 0.0178885 = -2.23607.

Part D: The *p* value for the one-tailed alternative hypothesis is the cumulative distribution function of the standard normal distribution at x = -2.23607, which is 0.01267365 (table look-up or spread-sheet function).

Part E: The *p* value for the two-tailed alternative hypothesis is twice the *p* value for the one-tailed alternative hypothesis = 2 × 0.01267365 = 0.0253473.

Part F: If the null hypothesis is true, the *z* value is normalized (subtract the mean and divide by the standard deviation), so it has a standard normal distribution, with a mean of zero and a variance of 1.

Part G: If the null hypothesis is not true, the *z* value does not have a standard normal distribution.

- The *z* value for testing the null hypothesis assumes a mean of μ_0.
- The true mean is μ'.

The denominator of the *z* value still uses the assumed proportion μ_0, but the computed *z* value for any sample is shifted by $(\mu' - \mu_0)$ / the denominator of the *z* value.

The expected value of the z value in this exercise is

$(0.75 - 0.80) / (0.80 \times (1 - 0.80) / 500)^{0.5} = -2.795085$

Question: The observed proportion in the sample is 76%, not 75%.

Answer: The observed proportion in the sample is distorted by random fluctuations. The problem says the true proportion is 75% and asks: "what is the expected value of the z value which assumes a proportion of 80%?"

Part H: The z value is (observed proportion $- p_0) / (p_0 (1 - p_0) / n)^{0.5}$.

The expected value of the observed proportion is p' and its variance is $p' \times (1 - p')$. The z value can be written as $k \times$ observed proportion $- k'$, where k and k' are constants.

The constant $k = [(p' (1 - p') / n) / (p_0 (1 - p_0) / n)]^{½}$.

The variance of $k \times Z - k' = k^2 \times$ the variance of Z (which is 1 for a standard normal distribution), so the variance of the z value (if the true proportion is p') = $(p' (1 - p') / n) / (p_0 (1 - p_0) / n)$.

Question: The division by n cancels in the numerator and the denominator.

Answer: We keep them here to remind us of the derivation; the textbook also keeps them.

The variance of the z value in this exercise is $(0.75 \times (1 - 0.75)) / (0.80 \times (1 - 0.80)) = 1.171875$

Part I: The probability of a Type II error when the true proportion is p' for the one-tailed test is

$$1 - \Phi[(p_0 - p' - z_\alpha \times (p_0 (1 - p_0) / n)^{0.5}) / ((p'(1 - p') / n)^{0.5})$$

We have $(p'(1 - p') / n)^{0.5} = 0.019364$ and $(p_0 (1 - p_0) / n)^{0.5} = 0.017889$, so the ratio for Φ is

$$(0.80 - 0.75 - 2.0537489 \times 0.017889) / 0.019364 = 0.6848$$

and

$$1 - \Phi(0.6848) = 0.2467$$

Part J: The probability of a Type II error when the true proportion is p' for the two-tailed test is

$$\Phi[(p_0 - p' + z_{\alpha/2} \times (p_0 (1 - p_0) / n)^{0.5}) / ((p'(1 - p') / n)^{0.5})$$

$$- \Phi[(p_0 - p' - z_{\alpha/2} \times (p_0 (1 - p_0) / n)^{0.5}) / ((p'(1 - p') / n)^{0.5})$$

We have $(p'(1 - p') / n)^{0.5} = 0.019364$ and $(p_0 (1 - p_0) / n)^{0.5} = 0.017889$, so

- the ratio for the first Φ is $(0.80 - 0.75 + 2.326348 \times 0.017889) / 0.019364 = 4.731$
- the ratio for the second Φ is $(0.80 - 0.75 - 2.326348 \times 0.017889) / 0.019364 = 0.433$

$$1 - \Phi[(p_0 - p' - z_{\alpha/2} \times (p_0 (1 - p_0) / n)^{0.5}) / ((p'(1 - p') / n)^{0.5})$$

and

$$\Phi(4.731) - \Phi(0.433) = 0.3325$$

Part K: The number of observations needed for a one-tailed test =

$$((z_\alpha \times (p_0 (1 - p_0))^{0.5} + z_\beta \times (p'(1 - p'))^{0.5}) / (p_0 - p'))^2$$

For $\alpha = 2\%$ and $\beta = 10\%$, this equals

$$((2.05375 \times (0.8 \times 0.2)^{0.5} + 1.28155 \times (0.75 \times 0.25)^{0.5}) / (0.75 - 0.80))^2 = 757.821$$

The next highest integer is 758.

Part L: The number of observations needed for a two-tailed test =

$$((z_{\alpha/2} \times (p_0 (1 - p_0))^{0.5} + z_\beta \times (p'(1 - p'))^{0.5}) / (p_0 - p'))^2$$

For $\alpha = 2\%$ and $\beta = 10\%$, this equals

$$((2.326349 \times (0.8 \times 0.2)^{0.5} + 1.28155 \times (0.75 \times 0.25)^{0.5}) / (0.75 - 0.80))^2 = 882.645$$

The next highest integer is 883.