MS Module 8: Two Population Proportions & Variances – practice problems

(The attached PDF file has better formatting.)

Exercise 8.1: Difference in population proportions

A study on a treatment group vs a control group shows

	treatment	*control*
observations	120	100
successes	72	50

The null hypothesis is H_0: $p_1 = p_2$, where p_1 and p_2 are the true proportions of success for the two groups.

A. What are the sample proportions of success in the treatment group and the control group?
B. What is the sample proportion of success in the combination of the two groups?
C. What is the sample difference in the proportions of success between the two groups?
D. What is the variance of the sample difference in the proportions of success between the two groups?
E. What is the standard deviation of the sample difference in the proportions of success?
F. What is the *z* statistic to test the null hypothesis H_0: $p_1 = p_2$?
G. What is the *p* value for this two-tailed *z* test of the null hypothesis H_0: $p_1 = p_2$?

Part A: The sample proportions of success in the treatment group and the control group are

- Treatment group: 72 / 120 = 60%
- Control group: 50 / 100 = 50%

Part B: The sample proportion of success in the combination of the two groups = 122 / 220 = 55.45455%.

	treatment	*control*	*combined*
observations	120	100	220
successes	72	50	122
sample proportion	0.6	0.5	0.554545

Part C: The sample difference in the proportions of success between the two groups is 60% – 50% = 10%.

Part D: The variance of the sample differences in the proportion of success between the two groups is

$$\hat{p} \times (1 - \hat{p}) \times (1/m + 1/n) = 0.554545 \times (1 - 0.554545) \times (1/120 + 1/100) = 0.004529.$$

Part E: The standard deviation is the square root of the variance = $0.004529^{0.5}$ = 0.067298.

Part F: The *z* statistic to test the null hypothesis H_0: $p_1 = p_2$ is the difference in the sample proportions divided by its standard deviation = 10% / 0.067298 = 1.485928.

Part G: The *p* value for this two-tailed *z* test of the null hypothesis H_0: $p_1 = p_2$ is 2 × CDF(–1.4859) = 0.1373.

Exercise 8.2: Type II errors and required sample sizes

The incidence of a disease in an untreated population is 2%. We test whether a vaccine reduces the incidence of the disease, using a 5% significance level. We want enough subjects so that if the incidence level of the disease with the vaccine is 1% or less, the probability of a Type II error is 10% or less.

A. What is z_α for this exercise?
B. What is z_β for this exercise?
C. How many subjects are needed for the given values of α and β?

Part A: This scenario is a one-tailed test for α = 5%, so we use z_α (not $z_{\alpha/2}$) = 1.644854.

Part B: The given β is 10%, so z_β = 1.281552.

Part C: For p_1 = 2% and p_2 = 1% and the given α and β, the number of subjects needed is

$n = [\, z_\alpha \sqrt{((p_1 + p_2)(q_1 + q_2)/2)} + z_\beta \sqrt{(p_1q_1 + p_2q_2)}\,]^2 / (p_1 - p_2)^2 =$

$(1.644854 \times ((2\%+1\%) \times (98\% + 99\%)/2)^{0.5} + 1.281552 \times (2\% \times 98\% + 1\% \times 99\%)^{0.5})^2 / (2\% - 1\%)^2 = 2{,}528.7$

Exercise 8.3: Confidence interval for difference in proportions

The observations and success for a treatment group and a control group are

	treatment	*control*
observations	120	100
successes	72	50

A. What is the sample difference in the probability of success between the treatment and control groups?
B. What is the sample variance of the difference in the probability of success between the two groups?
C. What is the sample standard deviation of this difference in the probability of success?
D. What is the 95% confidence interval for the null hypothesis that the probability of success is the same for the two groups?

Part A: The sample difference in the probability of success between the treatment and control groups is

72 / 120 – 50 / 100 = 0.60 – 0.50 = 0.10

Part B: The sample variance of the difference in the probability of success between the two groups is

0.60 × (1 – 0.60) / 120 + 0.50 × (1 – 0.50) / 100 = 0.0045

Question: To test the null hypothesis H_0: $p_1 = p_2$, we used the variance of the combined group:

$$\bar{p} \times \bar{q} \times (1/m + 1/n)$$

where $\bar{p} = (mp_1 + np_2) / (m + n)$ and $\bar{q} = (mq_1 + nq_2) / (m + n)$.

Answer: We use the variance of the combined group when we test the null hypothesis H_0: $p_1 = p_2$.

The confidence interval does not assume that $p_1 = p_2$ and we form an interval around the observed difference.

Part C: The sample standard deviation of the difference in the probability of success between the two groups is $0.0045^{0.5}$ = 0.067082

Part D: The *z* value for a two-sided 95% confidence interval is 1.959964. The 95% confidence interval is

- Lower bound: 0.1 – 1.959964 × 0.067082 = -0.031478
- Upper bound: 0.1 + 1.959964 × 0.067082 = 0.231478

Exercise 8.4: Difference of variances

Groups #1 and #2 are normally distributed.

- σ^{21} = the variance of Group #1
- σ^{22} = the variance of Group #2

The null hypothesis is H_0: $\sigma^{21} = \sigma^{22}$.

- A sample from Group #1 is {1, 3, 5, 7, 9, 11}.
- A sample from Group #2 is {12, 13, 14, 15, 16, 17, 18, 19}.

(To illustrate the procedure, these samples have uniform distributions with different variances, not normal distributions with the same variance.)

A. What are the variances of the samples from Group 1 and Group 2?
B. What is the ratio of the variances?
C. What is the distribution of this ratio of variances?
D. What is the *p* value for testing the null hypothesis?
E. What is the 95% confidence interval for the ratio of the variances of the two groups?
F. What is the 95% confidence interval for the ratio of the standard deviations of the two groups?

Part A: The mean of Group 1 is (1 + 3 + 5 + 7 + 9 + 11) / 6 = 6.00.

The sample variance of Group 1 is

$$((1-6)^2 + (3-6)^2 + (5-6)^2 + (7-6)^2 + (9-6)^2 + (11-6)^2) / (6-1) = 14.00$$

The mean of Group 2 is (12 + 13 + 14 + 15 + 16 + 17 + 18 + 19) / 8 = 15.50

The sample variance of Group 2 is

$$((12-15.5)^2 + (13-15.5)^2 + (14-15.5)^2 + (15-15.5)^2 + (16-15.5)^2 + (17-15.5)^2 + (18-15.5)^2 + (19-15.5)^2) / (8-1) = 6.00$$

Part B: The ratio of these variances is 14 / 6 = 2.333333.

Part C: If both groups are normally distributed and their variances are equal, the ratio of the sample variances has an *F* distribution with (m-1) and (n-1) degrees of freedom.

The textbook explains that if $X_1, \dots, X_m$ is a random sample from a normal distribution with variance σ^{21} and $Y_1, \dots, Y_n$ is another random sample (independent of the X_i's) from a normal distribution with variance σ^{22} and s^{21} and s^{22} are the two sample variances, then the random variable $F = (s^2_1 / \sigma^2_1) / (s^{22} / \sigma^{22})$ has an *F* distribution with degrees of freedom υ_1 = m-1 and υ_2 = n-1 (equation 10.8).

Part D: The *p* value for testing the null hypothesis H_0: $\sigma^{21} = \sigma^{22}$ is $F_{2.3333, 5, 7} = 0.14989$.

Question: Which of the two groups is the numerator of the ratio of the variances? The exercise here uses the variance of Group 1 as the numerator and the variance of Group 2 as the denominator. If we use the variance of Group 2 as the numerator and the variance of Group 1 as the denominator, the ratio of the variances is

6 / 14 = 0.428571, which is much different from 14 / 6 = 2.333333.

Answer: The critical values for the *F* distribution have the relation $F_{\alpha, s, t} = 1 / (F_{\alpha, t, s})$

Interchanging Group 1 with Group 2 replaces F by $1/F$ and interchanges the degrees of freedom s and t. The statistical tests (p values and confidence intervals) have the same results.

Part F: We look up (or compute) two critical F values for the 95% confidence interval ($\alpha = 0.05$):

- $F_{\alpha/2,\, s,\, t} = F_{0.025,\, 5,\, 7} = 5.285237$
- $F_{\alpha/2,\, t,\, s} = F_{0.025,\, 7,\, 5} = 6.853076$

The bounds for the 95% confidence interval divide or multiply the ratio of the sample variances by the critical F values:

- Lower bound: $[\ \sigma^{21} / \sigma^{22}\ ] \times 1 / F_{\alpha/2,\, s,\, t} = 2.333333 / 5.285237 = 0.441481$
- Upper bound: $[\ \sigma^{21} / \sigma^{22}\ ] \times F_{\alpha/2,\, t,\, s} = 2.333333 \times 6.853076 = 15.990508$

Part G: The bounds for the 95% confidence interval for the ratio of the standard deviations are the square roots of the bounds for the 95% confidence interval for the ratio of the variances:

- Lower bound: $0.441481^{0.5} = 0.664440$
- Upper bound: $15.990508^{0.5} = 3.998813$