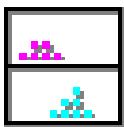


Solutions to Worktext Exercises



Chapter 10

Visualizing Two-Sample Hypothesis Tests

Basic Learning Exercises

1. The null hypothesis ($H_0: \mu_1 = \mu_2$) says that, on average, day and night students perform the same. The alternative hypothesis ($H_1: \mu_1 \neq \mu_2$) says that there is a difference in the performance of day and night students. Unless there is a prior reason to believe that either day or night students perform better, a two-tail test is appropriate. The null hypothesis is false because the scroll bar indicates a slight difference in means. However, such a small difference may be hard to detect.
2. Both populations are normal. The mean of the red population (day students) differs very slightly from the mean of the blue population (night students). Yes, the variances are equal, based on visual inspection of the distributions and on the position of the Difference in Variances scroll bar.
3. The red and blue samples overlap substantially, and their means are almost identical in most samples. The ranges and end points of the red and blue sampling distributions are similar. This supports the expectation that it will be difficult to detect such a slight difference in means. Because its sample size is slightly larger (26 versus 22), the sampling distribution for the night class (blue curve) has a smaller variance than the sampling distribution for the day class (red curve). This makes the sampling distribution of the night class taller and narrower than the sampling distribution of the day class.
4. a) If the null hypothesis (equal means) is true and if both populations are normal with known variances, the sampling distribution of the difference of means is exactly normal. b) No, sample size is irrelevant to the normality issue when the populations are normal and the variances are known. c) The critical values shown on the decision rule ($z = \pm 1.96$) are based on the standard normal distribution. They correspond to the chosen level of significance and do not vary from sample to sample. d) The test statistic varies because it depends on the sample estimates of the means that are subject to random variation.
5. a) Yes, the confidence intervals and the z-test always agree. b) The test statistic falls in the “acceptance” region (i.e., the confidence interval includes zero) almost every time (at least 9 out of 10 times). The decision would be to “accept” H_0 almost every time, even though H_0 is false. c) You cannot commit Type I error because H_0 is false, although you will commit Type II error almost every time. d) The test has low power in this example. No, this conclusion about power does not hold in general.
6. a) The true population means are 60.0 and 60.6. b) The difference is only 0.6 (about a 1% difference). c) The confidence intervals always agree.
7. The analysis reminds you that H_0 is false, and that (most likely) you committed Type II error (accepting a false null hypothesis). Given the difference in means, there is only about a 5% chance that you correctly rejected H_0 because the difference is so small.

Intermediate Learning Exercises

8. Experiment: 1 2 3 Total Avg. Power

Rejections: 6 4 8 18 0.06

The total is near 18 and the average power is about 0.06. Repetition helps average out the substantial random variation. The normal distribution is based on the assumption that H_0 is true. The histogram of test statistics is bell-shaped and slightly shifted relative to the hypothesized distribution, but the distribution and histogram overlap greatly.

9. Experiment: 1 2 3 Total Avg. Power

Rejections: 55 64 59 183 0.61

The sampling distributions overlap much less than before, so power should be much larger. The replication supports this prediction. The total should be near 183 and the average power should be about 0.61. The histogram is clearly shifted, but still overlaps the hypothesized normal distribution. Power is much greater than in exercise 8 because the difference of means is much greater.

10. Experiment: 1 2 3 Total Avg. Power

Rejections: 86 94 90 270 0.90

The sampling distributions become narrower and now barely overlap at all. The total should be near 270 and the average power should be about 0.90. Yes, the doubled sample sizes helped a lot (power has risen nearly 50% from 0.61 to 0.90). The histogram of test statistics is greatly shifted relative to the hypothesized normal distribution.

11. The t distribution is used in a test of two means when the variances are unknown.

12. The t-test assuming equal variances uses $df = n_1 + n_2 - 2 = 7 + 7 - 2 = 12$. Unless the sample sizes or the assumptions change, the degrees of freedom and critical values stay the same.

13. When variances are assumed equal, degrees of freedom are always $n_1 + n_2 - 2$. However, when variances are assumed unequal, degrees of freedom are adjusted because of the Behrens-Fisher problem. Adjusted degrees of freedom will be at least the smaller of $n_1 - 1$ or $n_2 - 1$, and will not exceed $n_1 + n_2 - 2$. Since Welch's formula depends on s_1^2 and s_2^2 (which vary from sample to sample) the adjusted degrees of freedom will vary.

14. The empirical Type I error should be 0.05 (results will range from about 0.01 to about 0.09). The histogram very closely resembles the hypothesized t distribution, as it should since the null hypothesis is true (equal means) and the variances are equal (as assumed).

15. Empirical Type I error still is about 0.05 (results may range from about 0.01 to 0.09). The histogram of test statistics still resembles the hypothesized t distribution. Type I error is apparently not much affected when unequal variances are assumed equal.

16. Empirical Type I error should be 0.05 (results will range from about 0.01 to about 0.09). The histogram closely resembles the hypothesized t distribution, as it should since the null hypothesis is true (equal means) and the assumed unequal variances are unequal.

17. The means are moderately unequal. The variances are very unequal. The variances are unknown but are (correctly) assumed unequal. The sample sizes are equal and are fairly small ($n_1 = 7$, $n_2 = 7$). The empirical power is about 0.16 (results will range from about 0.11 to 0.21). The histogram of test statistics is shifted in relation to the t distribution, and its shape may not be quite the same as the t-distribution.

18. The empirical power is about 0.63 (results will range from about 0.57 to 0.69). Increasing the sample sizes increased power substantially. Results are similar to exercise 10. Power increases as sample size increases if the difference in means is large enough to be detected.

19. $H_0: \sigma_1^2 = \sigma_2^2$ (population variances are the same) or $\sigma_1^2/\sigma_2^2 = 1$ (ratio is unity)

$H_1: \sigma_1^2 \neq \sigma_2^2$ (population variances are different) or $\sigma_1^2/\sigma_2^2 \neq 1$ (ratio isn't unity)

The true variance ratio is 2.25:1. It would seem likely that the difference in variance can be detected through sampling, although the sample sizes are rather small.

20. Numerator df	<u>9</u>	Left-tail critical value	<u>0.248</u>
Denominator df	<u>9</u>	Right-tail critical value	<u>4.026</u>

Since both populations are normal the sampling distribution for a ratio of sample variances is F under a true null hypothesis. The numerator degrees of freedom will be $n_1 - 1 = 10 - 1 = 9$ and the denominator degrees of freedom will be $n_2 - 1 = 10 - 1 = 9$.

21. The null hypothesis is accepted about 8 out of 10 times. No, the power of the test appears to be low. We know that H_0 is false but are unable to reject it most of the time because the difference in variances is not that great when using sample sizes of 10.

22. Experiment:	1	2	3	Total	Power
Rejections:	<u>23</u>	<u>16</u>	<u>19</u>	<u>58</u>	<u>0.19</u>

The average power is about 0.19, but there is considerable sample variation. Type I error is irrelevant because the null hypothesis is false. We cannot commit Type I error unless the null hypothesis is *true*.

23. It is shifted to the right and is much flatter than the hypothesized F distribution. It does not follow the F distribution because the null hypothesis is false.

Advanced Learning Exercises

24. Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	<u>0.01 to 0.09</u>	<u>0.01 to 0.09</u>	<u>0.01 to 0.09</u>	<u>0.05</u>
Skewed Right	<u>0.01 to 0.09</u>	<u>0.01 to 0.09</u>	<u>0.01 to 0.09</u>	<u>0.05</u>

The histogram resembles the t when sampling a normal population, but may seem very slightly right-skewed when sampling a skewed population (i.e., more rejections in the right tail than in the left). Type I error rates vary (0.05 ± 0.04), but are about the same regardless of the population shape. In terms of Type I error, the difference of sample means is robust to skewness, even for these fairly small samples.

25. Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	<u>0.01 to 0.09</u>	<u>0.01 to 0.09</u>	<u>0.01 to 0.09</u>	<u>0.05</u>
Skewed Right	<u>0.16 to 0.24</u>	<u>0.16 to 0.24</u>	<u>0.16 to 0.24</u>	<u>0.20</u>

When the population is skewed, the histogram of test statistics is flatter than the hypothesized F distribution. Type I error is much higher than the chosen level of significance $\alpha = 0.05$ (in this example, about four times greater). No, the test for two variances is not robust to skewness, in terms of Type I error. No, the large sample sizes didn't help. Yes, skewed populations are common, so it is a real problem.

26. The sampling distributions are fairly distinct, so we expect the test of two means to have reasonable power. No, unless the null hypothesis is true, Type I error cannot occur.

<u>27. Distribution</u>	<u>Experiment 1</u>	<u>Experiment 2</u>	<u>Experiment 3</u>	<u>Avg Power</u>
Normal	<u>0.40 to 0.57</u>	<u>0.40 to 0.57</u>	<u>0.40 to 0.57</u>	<u>0.50</u>
Skewed Right	<u>0.42 to 0.64</u>	<u>0.42 to 0.64</u>	<u>0.42 to 0.64</u>	<u>0.57</u>

Because the null hypothesis is false, the histogram is flatter than the t and is shifted substantially. The power estimates vary about ± 0.08 , but there is no evidence that skewness adversely affects power.

<u>28. Distribution</u>	<u>Experiment 1</u>	<u>Experiment 2</u>	<u>Experiment 3</u>	<u>Avg Power</u>
Normal	<u>0.15 to 0.25</u>	<u>0.15 to 0.25</u>	<u>0.15 to 0.25</u>	<u>0.20</u>
Skewed Right	<u>0.21 to 0.32</u>	<u>0.21 to 0.32</u>	<u>0.21 to 0.32</u>	<u>0.26</u>

The false null hypothesis makes the histogram of test statistics flatter than the hypothesized F distribution. When the populations are normal, power is much greater than Type I error. When the populations are skewed, there is little difference and so the probability of rejecting H_0 if it is true is about the same as if H_0 is false. Therefore, having a false H_0 does not increase the probability of rejecting H_0 . Hence, having the wrong population shape makes the test nearly useless.