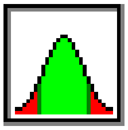
Solutions to Worktext Exercises



Chapter 9

Visualizing One-Sample Hypothesis Tests

Basic Learning Exercises

1. Hypothesized mean ____794.0____ True Mean ____794.0____
   The null hypothesis is true because the hypothesized and true means are the same.

2. Smallest sample observation __787__ Largest sample observation __802__
   Smallest sample mean __793__ Largest sample mean __796__
   Answers will vary but are likely to resemble the values shown here. The means are less variable than the sample items, as would be expected by the Central Limit Theorem (the standard deviation of the means is $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 3/\sqrt{15} = 0.775$ instead of $\sigma = 3$).

3. Minimum of X distribution __785__ Maximum of X distribution __803__
   Minimum of sampling dist. __792__ Maximum of sampling dist. __797__
   They agree in a general way. However, the distribution's maximum and minimum are likely to be more extreme than the samples obtained in 10 trials.

4. a) The test uses a normal distribution because the population variance is assumed known. b) The critical values are $z = \pm 1.96$. c) The critical values do not change from sample to sample. d) The test statistic does change from sample to sample.

5. Since the level of significance is $\alpha = 0.05$, we would expect 19 of the 20 samples to include the true value (i.e., 95% of the intervals). We would not be surprised to see 20 (i.e., 100%), but it would be quite unusual to see fewer than 18 (i.e., 90%).

6. Yes, they always agree.

7. Hypothesized mean ____490____ True Std. Dev. ____100____
   True mean ____510____ Sample size ____6____
   The null hypothesis is false because the hypothesized and true means are different.

8. a) When population variance is assumed unknown, the sampling distribution of the mean is Student's t with 5 degrees of freedom. Its critical values are ± 2.571. With known population variance, the sampling distribution of the mean is normal. Its critical values are ± 1.960. Yes, the critical values are substantially different. b) If the sample size were larger, the difference between Student's t and normal would be less noticeable.

9. a) Assuming *known* variance, $\sigma = 100$ is used in the confidence interval $\overline{X} \pm z \;\; \sigma/\sqrt{n}$. The interval width does not change from sample to sample (although the mean may change). b) Assuming *unknown* variances, the sample standard deviation s is used in each confidence interval $\overline{X} \pm t \;\; s/\sqrt{n}$. Because of sample variation, the estimated standard deviation (s) will vary from sample to sample (as will the mean), thereby affecting the confidence interval. c) In actual sampling, the population variance is generally unknown, although in some applications (like quality control) historic data may be available.

Intermediate Learning Exercises

10. Experiment:

Experiment:	1	2	3	4	5	6	7	8	9	10
Rejections:	4	4	3	3	4	10	4	5	7	2

The mean should be near 5 (it is 4.6 in this illustrative set of 10 experiments). The range is approximately 2 to 10 rejections in 100 samples. With $\alpha = 0.05$ we would expect about 5 rejections in 100 since the null hypothesis is true, but obviously there is considerable variation around this expectation. This is due to the nature of random sampling. Yes, many observers would expect more stability in such a large number of repetitions.

11. If the manufacturer stops the filling process to adjust the settings every time a 95% confidence interval fails to include 794, then unnecessary stoppages will occur about 5% of the time (since the true mean is actually 794 grams). Lowering the level of significance (say to α = 0.01) would result in fewer unnecessary stoppages but would also make it less likely that real problems would be detected.

12. No. The preponderance of samples indicate that the null hypothesis is true ($\mu = 794$).

13. The histogram of Z test statistics is bell-shaped and correctly centered under the hypothesized normal sampling distribution. The number in each tail will approximate the red area. The Analysis of Experiment window says that from 1 to 4 rejections occur in each tail (average of 2.5) for an empirical Type I error of about 0.05.

14. None of the confidence intervals encloses the true mean (they are colored red and lie to the left of $\mu = 794$). The empirical power (note that H_0 is false) is basically 100%. The histogram is still normal, but is shifted clear to the left of the hypothesized sampling distribution. It is easy to detect a difference of 4 grams with this sample size and true standard deviation. No, it is not safe to generalize. If the sample size were smaller or if the standard deviation were larger, a difference of 4 grams could be harder to detect.

15. The true sampling distribution lies to the left and does not overlap the hypothesized sampling distribution, so the power of the test is extremely high. The standard error of the mean is only $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 3/\sqrt{15} = 0.775$, so 4 grams is a large difference.

16. Although the true sampling distribution lies slightly to the left of the hypothesized sampling distribution, they overlap considerably. Therefore, the difference of 4 grams will be much more difficult to detect and the hypothesis test will have much lower power than in the previous example with $\sigma = 0.05$.

17. a) The null hypothesis will be rejected about 29 times (results will range from 22 to 36). This is the same as saying the confidence intervals will fail to include $\mu = 794$ about 29% of the time. b) The approximate power of the test is 0.29 (ratio of times H_0 is correctly rejected to the number of replications). c) Type I error is not relevant because H_0 is false. The histogram of test statistics is clearly shifted to the left of the hypothesized sampling distribution, with about half of its area lying inside the left-tail rejection region.

18. The observed power is about 0.07 (answers will vary from about 0.05 to about 0.09), which implies that we will make a correct decision to reject the null hypothesis (which is indeed false) only about 7 percent of the time (and commit Type II error about 93 percent of the time). The histogram of test statistics is slightly shifted to the right of the hypothesized distribution, but not enough to give us confidence in detecting the false H_0.

19. The observed power improves to about 0.16 (answers will vary from about 0.15 to about 0.18) but is still low. The histogram is noticeably shifted to the right, but still overlaps the hypothesized sampling distribution. Yes, larger sample size increased the power of the test, but it is too low to be of great practical value.

20. No. Even though the true mean is actually higher than 490, the majority of samples are unable to detect this fact.

21. Hypothesized variance 0.0100 True variance 0.0169
Sample size 10 Test type Left tail
The null hypothesis is false because the hypothesized and true variances differ.

22. Degrees of freedom 9 Critical value 16.919
If the population is normal, the sampling distribution for a sample variance is chi-square with DF = n – 1 = 10 – 1 = 9. A right-tail test for excessive variation is appropriate because no one will complain if variation is less than expected (the cups will be filled with greater accuracy). The mean is of lesser importance in this scenario because the decision maker is concerned with variance about the mean.

23. The confidence interval is open-ended in the same direction as the alternate hypothesis H_1 because it is a right-tailed test. Yes, it applies to any one-tailed test.

24. The outcomes of the tests (using either approach) are uncertain. The null hypothesis is false, but will be rejected only about 1/3 of the time. The approaches always agree.

25. Empirical power is about 0.34 (in replication experiments, power estimates will range roughly from 0.31 to 0.38). Type I error is irrelevant because the null hypothesis is false. We cannot commit Type I error unless the null hypothesis is *true*.

26. Empirical power is about 0.34 (in replication experiments, power estimates will range roughly from 0.31 to 0.38).

27. a) No, the empirical power should be about the same in both experiments. b) The larger number of replications would presumably yield more stable estimates of empirical power. Otherwise, it doesn't matter. c) The confidence interval display is already crowded with 500 replications. With 1000 replications, the confidence intervals would be so dense as to be virtually illegible.

Advanced Learning Exercises

28.

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	0.02 to 0.08	0.02 to 0.08	0.02 to 0.08	0.05
Uniform	0.02 to 0.08	0.02 to 0.08	0.02 to 0.08	0.05
Skewed Right	0.04 to 0.10	0.04 to 0.10	0.04 to 0.10	0.06
Very Skewed R	0.04 to 0.10	0.04 to 0.10	0.04 to 0.10	0.06

The histogram is normally distributed when sampling a normal or uniform population, but is noticeably right-skewed when sampling either a right-skewed or very right-skewed population. Average Type I error is 0.05 ± 0.03 for the normal and uniform distributions but is about 0.01 higher for the skewed distributions.

29.

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	0.02 to 0.08	0.02 to 0.08	0.02 to 0.08	0.05
Uniform	0.02 to 0.08	0.02 to 0.08	0.02 to 0.08	0.05
Skewed Right	0.02 to 0.09	0.02 to 0.09	0.02 to 0.09	0.05
Very Skewed R	0.02 to 0.09	0.02 to 0.09	0.02 to 0.09	0.05

The larger sample size makes the histograms more nearly normally distributed when the population is skewed or very skewed, as predicted by the Central Limit Theorem. Average Type I error is nearly the same regardless of the population shape (usually 0.05 ± 0.03) but is slightly higher for the skewed populations. In terms of Type I error, the sample mean appears robust to non-normality. Sample size has a slight effect on variation in empirical Type I error between experiments.

30. The standard error of the mean $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 10/\sqrt{4} = 5$ is fairly large in relation to the difference between the hypothesized mean ($\mu_0 = 100$) and true mean ($\mu = 96$), so the test will have low power in detecting the false H_0.

31.

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Power
Normal	0.07 to 0.17	0.07 to 0.17	0.07 to 0.17	0.12
Uniform	0.07 to 0.17	0.07 to 0.17	0.07 to 0.17	0.13
Skewed Right	0.07 to 0.17	0.07 to 0.17	0.07 to 0.17	0.11
Very Skewed R	0.07 to 0.17	0.07 to 0.17	0.07 to 0.17	0.11

Regardless of the population, the histogram is slightly shifted to the left of the true sampling distribution. The average power is about the same (0.12 ± 0.05), but random variation in experiments may suggest slight differences to some observers.

32.

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Power
Normal	0.46 to 0.59	0.46 to 0.59	0.46 to 0.59	0.52
Uniform	0.46 to 0.60	0.46 to 0.60	0.46 to 0.60	0.53
Skewed Right	0.45 to 0.61	0.45 to 0.61	0.45 to 0.61	0.54
Very Skewed R	0.45 to 0.61	0.45 to 0.61	0.45 to 0.61	0.54

Regardless of the population, the histogram is clearly shifted to the left of the true sampling distribution. Increasing the sample size has dramatically improved power (about a fourfold increase) because the standard error $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 10/\sqrt{25} = 2$ is now smaller relative to the difference between the hypothesized mean ($\mu_0 = 100$) and true mean ($\mu = 96$). The false H_0 is therefore more easily detected. Average power is similar for the various populations, so in terms of power the mean is apparently robust to non-normality. Variation is about ± 0.06 among the various experiments.

33.

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	0.02 to 0.07	0.02 to 0.07	0.02 to 0.07	0.05
Skewed Right	0.12 to 0.22	0.12 to 0.22	0.12 to 0.22	0.17
Very Skewed R	0.24 to 0.38	0.24 to 0.38	0.24 to 0.38	0.30

When the population is skewed, Type I error is much higher than the chosen level of significance $\alpha = 0.05$ (in this example, more than three times greater for a skewed population and about six times greater for a very skewed population). No, the large sample size didn't help. No, the direction of skewness wouldn't matter.

34. The histogram closely resembles a chi-square distribution when sampling a normal population, as it should when the null hypothesis is true. With a skewed or very right-skewed population, the histogram is flattened and its tails extend well beyond the critical values. This explains why Type I error was too high in the previous exercise (because the tails of the histogram were too long).

35. A skewed population's long tail results in erratic (albeit unbiased) estimates for the mean and variance, especially in small samples. The likelihood of extreme values in the sample makes tests of the variance unreliable. The mean is more stable, as has been seen in previous exercises. Tests for variance are *not* robust to non-normality, whereas tests for a mean *are* robust to non-normality.

36.

Sample Size	Experiment 1	Experiment 2	Average
n = 6	0.16 to 0.25	0.16 to 0.25	0.21
n = 24	0.38 to 0.48	0.38 to 0.48	0.43
n = 96	0.86 to 0.92	0.86 to 0.92	0.89

Power roughly doubles each time you quadruple the sample size. It would be reckless to generalize the conclusion about the *magnitude* of the effect of sample size on power, but it is safe to say that greater sample size implies greater power (other things equal). No, this ratio of sample variance to hypothesized variance (150 to 100) is not very extreme, statistically speaking (which may surprise some observers). The proof is that, in this example, it is difficult to reject H_0 reliably until the sample size is quite large.