ICS Part-II Statistics — Punjab Board

CHAPTER 13

Hypothesis Testing

Statistics • 2nd Year • ICS • Punjab Board • Free PDF Notes

Chapter 13: Hypothesis Testing

Chapter 12 answered the question 'what is a plausible range of values for an unknown parameter?' through confidence intervals. This chapter answers a different but closely related question: 'is a specific claimed value of that parameter believable, given what the sample shows?' This is hypothesis testing -- the second and last major branch of statistical inference, and by far the most heavily used tool in applied statistics, business decision-making, and scientific research. It is also the largest chapter in the book (70 pages), because it methodically works through the same battery of situations covered in Chapter 12 -- mean, proportion, difference of two means (independent and paired/dependent samples), difference of two proportions -- but now framed as formal decision procedures rather than interval estimates.

The chapter builds a single, disciplined framework -- state hypotheses, choose a significance level, pick a test statistic, define a rejection region, compute the observed value, and draw a conclusion -- and then applies that exact same framework, unchanged, to every situation that follows. Master the framework once in section 13.1, and the rest of the chapter is just plugging different test statistics (Z or T) into the same seven-step procedure.

Learning Objectives

- Define statistical hypothesis, null hypothesis, and alternative hypothesis, and correctly formulate them from a verbal claim
- Distinguish simple from composite hypotheses, and one-tailed from two-tailed tests
- Define Type-I and Type-II errors, and the level of significance and level of confidence
- List and apply the seven standard steps of a hypothesis test
- Test hypotheses about a population mean using the Z and T test statistics, choosing correctly between them
- Test hypotheses about a population proportion, with and without the continuity correction

- Test hypotheses about the difference between two population means for independent samples (known variance, pooled t, large-sample Z)
- Test hypotheses about the difference between two population means for dependent (paired/matched) samples
- Test hypotheses about the difference between two population proportions

Key Concepts

13.1 The Elements of a Test of Hypothesis

A statistical hypothesis is an assertion or conjecture about the distribution of one or more random variables -- typically phrased quantitatively in terms of a population parameter (e.g., mu=25, pi=0.4). Hypothesis testing is the procedure used to determine whether such an assumption is supported by an observed random sample. The null hypothesis, H0, is the hypothesis tested for possible rejection under the assumption that it is true -- it is always a statement of 'no effect' or 'the status quo' (a coin is unbiased, a drug is ineffective, there is no difference between two methods). The alternative hypothesis, H1, is the hypothesis we are willing to accept if H0 is rejected -- it is often called the research hypothesis, because it expresses what the researcher actually believes and is trying to support with evidence.

Formulating H0 and H1 correctly follows a fixed procedure: first state the alternative hypothesis as whatever the researcher is trying to find support for (an inequality: 'less than', 'greater than', or 'not equal to'), then state the null hypothesis as the complementary claim with an equality sign built in. The three matched pairs are: H0: theta>=theta0 vs H1: theta<theta0 (left-tailed); H0: theta<=theta0 vs H1: theta>theta0 (right-tailed); H0: theta=theta0 vs H1: theta!=theta0 (two-tailed). A simple hypothesis completely specifies the population distribution (both its functional form and every parameter value); a composite hypothesis does not (e.g., specifying only the mean while leaving the variance unspecified or as an inequality).

The test statistic is the sample statistic used to decide whether to reject H0 -- commonly Z, T, chi-square, or F. The rejection (critical) region is the set of test-statistic values for which H0 is rejected; the nonrejection region is the complementary set; the critical values are the boundary points separating them.

A two-tailed test places the rejection region equally in both tails of the sampling distribution (used when H1 is theta!=theta0); a one-tailed test places it entirely in one tail -- a left-tailed test when H1 is theta<theta0 (looking for a definite decrease), and a right-tailed test when H1 is theta>theta0 (looking for a definite increase).

Because a decision is being made from incomplete (sample) information, two kinds of error are possible. A Type-I error is rejecting H0 when H0 is actually true. A Type-II error is accepting (not rejecting) H0 when H1 is actually true. The probability of a Type-I error is denoted alpha and is called the level of significance (also the size of the critical region/test) -- a small, pre-assigned value such as 0.05 or 0.01, fixed BEFORE the sample is drawn so the results cannot bias the choice. The probability of a Type-II error is denoted beta. The level of confidence, 1-alpha, is the probability of correctly accepting a true H0. A hypothesis test concludes either by rejecting H0 in favour of H1 (if the test statistic falls in the rejection region), or by failing to reject H0 (there is insufficient evidence to reject it -- this is NOT the same as proving H0 true).

13.1.25 The Seven Steps of a Hypothesis Test

Every hypothesis test in this chapter follows the identical disciplined sequence. Before looking at the sample: (1) identify the population and state the conditions required for the test's validity; (2) formulate and state H0 and H1; (3) decide and specify the level of significance, alpha; (4) select the appropriate test statistic and its sampling distribution assuming H0 is true; (5) find the critical value(s) of the test statistic for the chosen alpha; (6) establish the rejection region; (7) state the decision rule (reject H0 if the test statistic falls in the rejection region, otherwise do not reject). After the sample is collected: (8) calculate the observed value of the test statistic from the sample; (9) draw the conclusion -- reject or do not reject H0 -- and state it in plain, practical terms.

13.2 Test of Hypothesis About a Population Mean, mu

The three possible hypothesis pairs about mu (against a hypothesised value mu0) are H0: mu>=mu0 vs H1: mu<mu0; H0: mu<=mu0 vs H1: mu>mu0; and H0: mu=mu0 vs H1: mu!=mu0. The choice of test statistic mirrors Chapter 12's confidence-interval logic exactly: when the population is normal with KNOWN

variance sigma^2 (any sample size), use Z = (X-bar - mu0)/(sigma/sqrt(n)), which is exactly standard normal under H0. When the population is normal with UNKNOWN variance and the sample is SMALL (n<=30), use T = (X-bar - mu0)/(S-hat/sqrt(n)), which follows a t-distribution with v=n-1 degrees of freedom under H0. When the sample is LARGE (n>30), the Central Limit Theorem justifies using Z = (X-bar - mu0)/(S-hat/sqrt(n)) as an approximation regardless of the population's shape or whether sigma is known.

In every case the mechanics are the same: compute the observed value of the test statistic from the sample, compare it against the critical value(s) that mark off the rejection region at the chosen alpha, and reject H0 if the observed value falls in that region.

13.3 Test of Hypothesis About a Population Proportion, pi

For a hypothesised proportion pi0, the three hypothesis pairs are analogous to those for the mean. Since the sampling distribution of the sample proportion P=X/n is (for large n) approximately normal with mean pi0 and variance pi0(1-pi0)/n, the test statistic is Z = (P - pi0)/sqrt[pi0(1-pi0)/n], or equivalently in terms of the count X: Z = (X - n.pi0)/sqrt[n.pi0(1-pi0)]. For extra precision, a continuity correction can be applied (since X is discrete but is being approximated by a continuous normal distribution): Z = [(X +/- 0.5) - n.pi0]/sqrt[n.pi0(1-pi0)] -- using a plus sign when X<n.pi0 (or P<pi0) and a minus sign when X>n.pi0 (or P>pi0).

13.4 Difference Between Two Means -- Independent Samples

For comparing mu1 and mu2 via independent samples, against a hypothesised difference delta0, three cases parallel Chapter 12 exactly. Known variances (any sample size, normal populations): Z = [(X-bar-1 - X-bar-2) - delta0] / sqrt[sigma1^2/n1 + sigma2^2/n2], exactly standard normal under H0. Small samples (n1,n2<=30), normal populations, equal but unknown variance: first pool the two sample variances into Sp^2 = [(n1-1).S-hat-1^2 + (n2-1).S-hat-2^2]/(n1+n2-2), then use T = [(X-bar-1 - X-bar-2) - delta0] / [Sp.sqrt(1/n1 + 1/n2)], which follows a t-distribution with v=n1+n2-2 degrees of freedom. Large samples (n1,n2>30), any population shape: the CLT justifies Z = [(X-bar-1 - X-bar-2) - delta0] / sqrt[S-hat-1^2/n1 + S-hat-2^2/n2] as an approximation, with no assumption of equal variances required.

13.5 Difference Between Two Means -- Dependent (Paired) Samples

When observations naturally come in matched pairs (before/after measurements on the same subject, or two products tested on the same set of items), the two samples are dependent rather than independent -- pairing lets each subject act as its own control, filtering out subject-to-subject variability that would otherwise inflate the noise in an independent-samples comparison. For each pair, define the difference D = Y - X (after minus before, or Candidate-A minus Candidate-B, etc.). The population of these differences has mean mu_D = mu2 - mu1 and its own variance sigma_D^2 -- and testing mu_D reduces the two-sample problem to an ordinary ONE-sample test on the differences, using exactly the machinery from section 13.2.

Given n paired differences d1, d2, ..., dn with mean d-bar = (sum d_i)/n and standard deviation s-hat_D = sqrt{[sum(d_i - d-bar)^2]/(n-1)}, the test statistic for H0: mu_D = delta0 (delta0 is usually 0, i.e. 'no difference') is T = (D-bar - delta0)/(s-hat_D/sqrt(n)), following a t-distribution with v=n-1 degrees of freedom. Two assumptions are required: the n differences must be a random sample from the population of differences, and that population of differences must be (approximately) normally distributed -- though this normality assumption can be relaxed for large n (>30) by the Central Limit Theorem, exactly as with a single-sample mean test.

13.6 Difference Between Two Proportions

For comparing pi1 and pi2 from two independent samples (sizes n1, n2, sample proportions P1=X1/n1, P2=X2/n2), two distinct situations arise. To test the GENERAL hypothesis H0: pi1-pi2=delta0 for some non-zero delta0, use Z = [(P1-P2) - delta0] / sqrt[P1(1-P1)/n1 + P2(1-P2)/n2], using the two sample proportions separately in the standard error. To test specifically whether the two populations have the SAME proportion of successes, H0: pi1=pi2 (i.e. delta0=0), it is more efficient to POOL the two samples into a single combined estimate of the common proportion, pi-hat = (X1+X2)/(n1+n2) = (n1.P1+n2.P2)/(n1+n2), and use Z = (P1-P2) / sqrt{pi-hat(1-pi-hat).(1/n1 + 1/n2)} -- exactly mirroring the pooled-variance logic used for two means with equal variance in section 13.4.

Important Definitions

What is a null hypothesis (H0)?: A statement of 'no effect' or 'the status quo' about a population parameter, tested for possible rejection under the assumption that it is true.

What is an alternative hypothesis (H1)?: The hypothesis accepted when H0 is rejected; it expresses the research claim the investigator is trying to support with sample evidence.

What is a Type-I error?: Rejecting the null hypothesis H0 when H0 is actually true; its probability is denoted alpha, the level of significance.

What is a Type-II error?: Failing to reject (accepting) the null hypothesis H0 when the alternative hypothesis H1 is actually true; its probability is denoted beta.

What is the level of significance?: The maximum probability of a Type-I error the researcher is willing to risk, denoted alpha; a small, pre-assigned value (e.g., 0.05 or 0.01) fixed before sampling.

What is the rejection (critical) region?: The set of values of the test statistic for which H0 is rejected in favour of H1.

What is the difference between a one-tailed and a two-tailed test?: A one-tailed test places the entire rejection region in one tail (used when H1 specifies a direction, < or >); a two-tailed test splits the rejection region equally between both tails (used when H1 is !=).

What is a simple hypothesis versus a composite hypothesis?: A simple hypothesis completely specifies the population distribution (functional form and all parameter values); a composite hypothesis leaves at least one parameter unspecified or given as an inequality.

Why are dependent (paired) samples used instead of independent samples?: Pairing lets each subject act as its own control, removing subject-to-subject variability from the comparison and increasing the power to detect a real difference, especially with small samples.

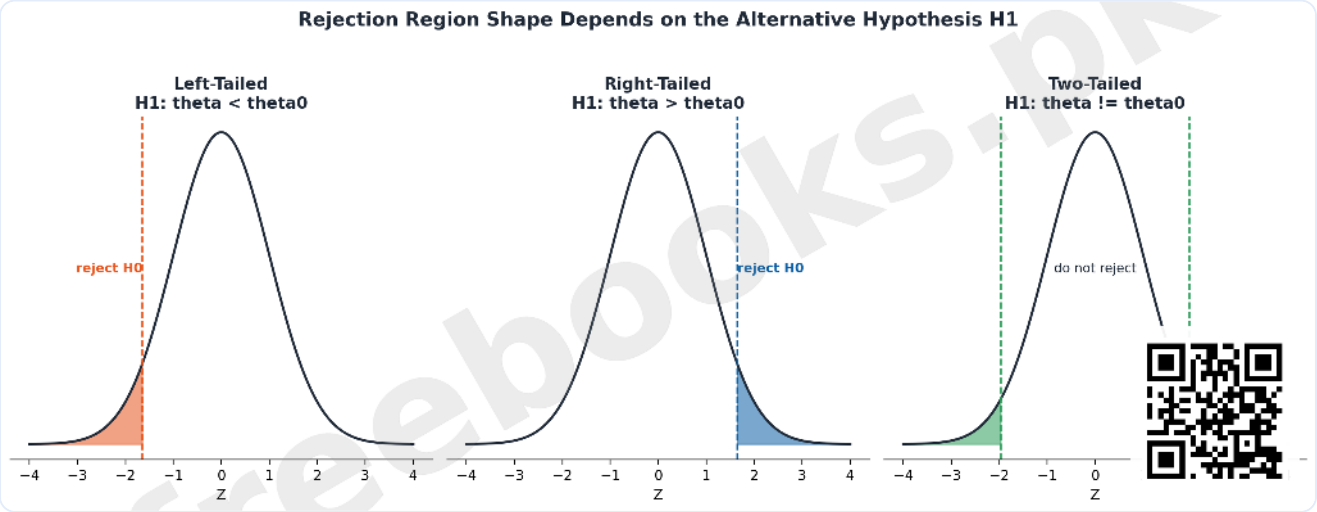
What is the pooled estimate of a common proportion, pi-hat, used for?: To test H0: pi1=pi2 (that two independent populations have the same proportion of successes), by combining both samples' successes into one estimate, pi-hat=(X1+X2)/(n1+n2), used in the standard error of P1-P2.

Key Facts and Relations

Topic	Key Fact / Relation
Test statistic for mu, sigma known / large n	Z = (X-bar - mu0) / (sigma/sqrt(n)) [or S-hat/sqrt(n) if sigma unknown, large n]
Test statistic for mu, small n, normal, sigma unknown	T = (X-bar - mu0) / (S-hat/sqrt(n)), v = n-1
Test statistic for proportion pi	Z = (P - pi0) / sqrt[pi0(1-pi0)/n]
Continuity-corrected Z for proportion	Z = [(X +/- 0.5) - n.pi0] / sqrt[n.pi0(1-pi0)]
Difference of means, known variances	Z = [(Xbar1-Xbar2) - delta0] / sqrt[sigma1^2/n1 + sigma2^2/n2]
Difference of means, small n, pooled variance	T = [(Xbar1-Xbar2) - delta0] / [Sp.sqrt(1/n1+1/n2)], v=n1+n2-2
Pooled variance, Sp^2	Sp^2 = [(n1-1)S1^2 + (n2-1)S2^2] / (n1+n2-2)
Difference of means, large samples	Z = [(Xbar1-Xbar2) - delta0] / sqrt[S1^2/n1 + S2^2/n2]
Paired-sample test (mu_D)	T = (D-bar - delta0) / (S_D-hat/sqrt(n)), v = n-1; D-bar = (sum d_i)/n
Difference of proportions, general delta0	Z = [(P1-P2) - delta0] / sqrt[P1(1-P1)/n1 + P2(1-P2)/n2]
Difference of proportions, testing pi1=pi2 (pooled)	Z = (P1-P2) / sqrt{pi-hat(1-pi-hat)(1/n1+1/n2)}, pi-hat=(X1+X2)/(n1+n2)
Error probabilities	alpha = P(reject H0 | H0 true) = P(Type-I); beta = P(accept H0 | H1 true) = P(Type-II)

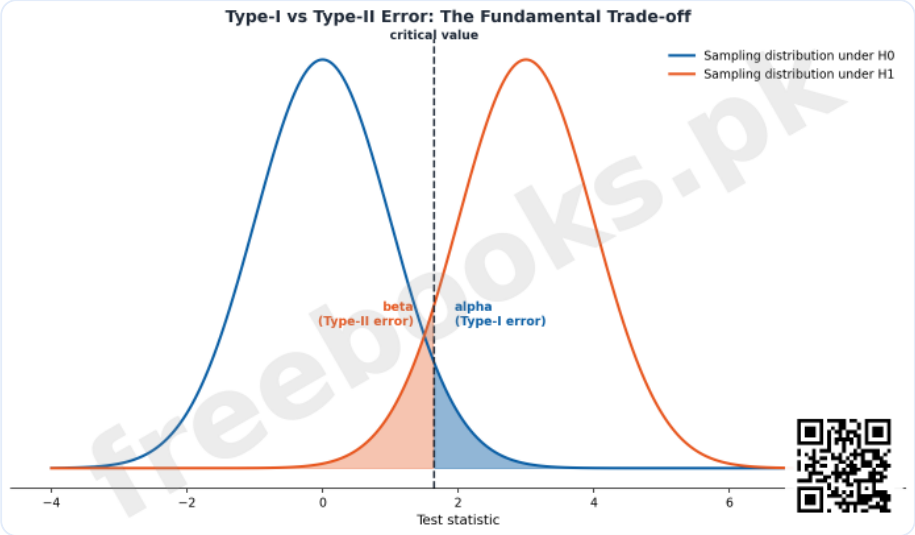
Diagrams

One-Tailed vs Two-Tailed Rejection Regions: Three standard normal curves side by side showing the shaded rejection region for a left-tailed test (Z<-z_alpha), a right-tailed test (Z>z_alpha), and a two-tailed test (split equally in both tails, Z<-z_alpha/2 or Z>z_alpha/2), illustrating how the shape of H1 determines where the critical region is placed



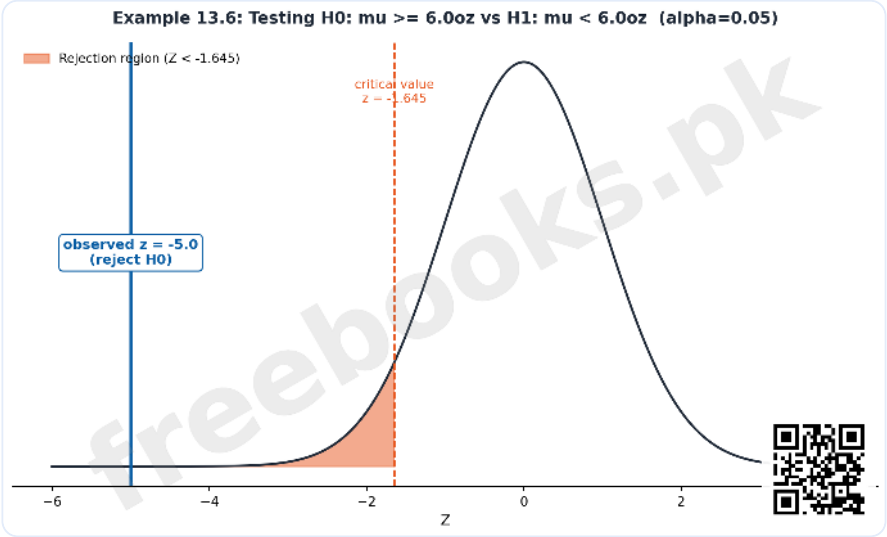
One-Tailed vs Two-Tailed Rejection Regions

Type-I and Type-II Error Trade-off: A two-panel diagram showing the sampling distribution of the test statistic under H0 and under H1 side by side (overlapping), with the alpha region (Type-I error, area under the H0 curve past the critical value) and the beta region (Type-II error, area under the H1 curve on the non-rejection side of the critical value) shaded and labelled, illustrating the fundamental trade-off between the two error types



Type-I and Type-II Error Trade-off

Worked Z-Test Example: Coffee Jar Weight: A standard normal curve for Example 13.6 (H0: mu>=6.0oz vs H1: mu<6.0oz, alpha=0.05) with the left-tail rejection region (Z<-1.645) shaded and the observed test statistic z=-5.0 plotted well inside the rejection region, visually confirming the decision to reject H0



Worked Z-Test Example: Coffee Jar Weight

Short Questions & Answers

Distinguish between the null hypothesis and the alternative hypothesis.

The null hypothesis (H0) is a statement of 'no effect' or the status quo, tested for possible rejection; the alternative hypothesis (H1) is the research claim accepted only if H0 is rejected, and expresses what the researcher is trying to demonstrate.

What is the difference between a Type-I error and a Type-II error?

A Type-I error rejects a true H0 (probability alpha); a Type-II error fails to reject a false H0, i.e., accepts H0 when H1 is actually true (probability beta).

Why is the level of significance, alpha, fixed BEFORE the sample is collected?

So that the chosen risk level cannot be influenced or biased by the actual sample results -- fixing alpha in advance keeps the test objective.

When should a one-tailed test be used instead of a two-tailed test?

A one-tailed test is used when the alternative hypothesis specifies a direction (a definite increase or decrease, i.e. H1 uses < or >); a two-tailed test is used when H1 simply states 'not equal to', with no specified direction.

Why does testing paired (dependent) samples reduce to a one-sample t-test on the differences?

Because each pair's difference D=Y-X collapses two measurements into a single number per subject; testing mu_D=mu2-mu1 then uses exactly the same one-sample machinery as testing an ordinary population mean, just applied to the sample of differences.

Why is a pooled estimate of variance (or proportion) needed for small-sample two-group tests?

Because two separate sample estimates (S1^2 and S2^2, or P1 and P2) of the same underlying population quantity will differ due to sampling error alone; pooling combines information from both samples into a single, more reliable estimate assumed common to both populations.

Long Questions & Answers

Explain the complete logical structure of a hypothesis test -- the null and alternative hypotheses, Type-I and Type-II errors, the level of significance, and the seven-step testing procedure -- using a concrete example to illustrate each concept.

Hypothesis testing provides a formal, disciplined framework for using sample evidence to decide between two competing claims about an unknown population parameter, and understanding this framework properly requires working carefully

through each of its interlocking pieces in turn, since they all depend on one another. Every hypothesis test begins with two competing statements about the population parameter under investigation. The null hypothesis, always denoted H0, is a statement of 'no effect' or 'the status quo' -- it is the specific claim that is put on trial and tested for possible rejection, under the working assumption that it is true unless the sample evidence strongly contradicts it. Typical null hypotheses include claims like 'this coin is unbiased,' 'this new drug is no more effective than the existing one,' or 'there is no difference between two production methods.' The alternative hypothesis, always denoted H1, is the competing claim that the researcher actually suspects or hopes to demonstrate -- it is often called the research hypothesis precisely because it usually expresses what the investigator set out to prove in the first place. Crucially, these two hypotheses must be formulated as an exhaustive, mutually exclusive pair covering every possible true value of the parameter, and the standard convention is to build the equality sign into H0 (using >=, <=, or =) while H1 takes the strict, direction-expressing complement (<, >, or !=). Because any decision reached from a hypothesis test is based on incomplete information, drawn from only a sample rather than the entire population, two distinct kinds of mistake are always possible, and distinguishing them carefully is essential. A Type-I error occurs when the null hypothesis is rejected even though it was actually true -- in effect, a false alarm, concluding an effect exists when it doesn't. A Type-II error occurs when the null hypothesis is NOT rejected even though the alternative hypothesis was actually true -- a missed detection, failing to find an effect that genuinely exists. These two kinds of error are asymmetric in an important sense: a hypothesis test as constructed in this chapter always fixes a small, pre-chosen risk of Type-I error, denoted alpha and called the level of significance, chosen deliberately BEFORE any sample data is even collected, specifically so this choice cannot be influenced or biased by whatever the data eventually turn out to show. The complementary probability, 1-alpha, is called the level of confidence, representing the probability of correctly failing to reject a true null hypothesis. Once H0, H1, and alpha have been settled, an appropriate test statistic must be selected -- a sample-based quantity, such as Z or T, whose exact sampling distribution under the assumption that H0 is true is known and can be looked up. Depending on the specific shape of H1, the rejection region -- the range of test-statistic values that would lead to rejecting H0 -- is placed either entirely in one

tail of that sampling distribution (a one-tailed test, used when H1 specifies a direction such as 'less than' or 'greater than') or split equally between both tails (a two-tailed test, used when H1 simply asserts 'not equal to', with no particular direction specified in advance). To make this entire framework fully concrete, consider testing whether a coffee-packing machine, claimed to fill jars with a mean of 6.0 ounces, might actually be underfilling them. Since the concern is specifically about UNDERfilling, the research (alternative) hypothesis is H1: mu<6.0, making this a one-sided, left-tailed test, with the corresponding null hypothesis H0: mu>=6.0. Suppose the significance level is set at alpha=0.05 before any sample is drawn, and a random sample of 100 jars is then weighed. Since the population standard deviation is assumed known here, the appropriate test statistic is Z=(X-bar-mu0)/(sigma/sqrt(n)), which follows a standard normal distribution if H0 is true. Looking up the critical value for a left-tailed test at alpha=0.05 gives z_alpha=-1.645, so the rejection region is Z<-1.645: any observed z-value more negative than -1.645 leads to rejecting H0. If the observed sample mean turns out to be 5.9 ounces, the observed test statistic computes to z=(5.9-6.0)/(0.2/sqrt(100))=-5.0, and because -5.0 is far more negative than the critical value of -1.645, it falls squarely inside the rejection region, so H0 is rejected: the evidence supports concluding the machine really is underfilling the jars, at the 5% significance level. If a Type-I error were made here, it would mean concluding the machine underfills when it actually doesn't; if a Type-II error had instead been made (by failing to reject H0 when the machine truly was underfilling), the company would continue believing everything is fine when customers are, in fact, being shortchanged.

Compare and contrast the test of the difference between two means for independent samples versus dependent (paired) samples, explaining when each design is appropriate, how each test statistic is constructed, and why pairing can be advantageous.

Comparing two populations' means is one of the most common tasks in applied statistics -- deciding whether a new fertilizer outperforms an old one, whether two machines fill containers to the same average weight, or whether a training programme actually improves worker productivity -- and the correct testing procedure depends critically on HOW the two samples being compared were actually collected, specifically on whether they are independent of one another

or deliberately paired (dependent). Two samples are independent if the selection of observations from one population has no bearing whatsoever on the selection of observations from the other population -- for instance, drawing 25 items from one machine's output and, entirely separately, drawing 36 items from a different machine's output. Two samples are instead dependent, or paired/matched, if each observation in one sample is deliberately linked to a specific, corresponding observation in the other sample, forming genuine pairs -- the classic example being the same group of individuals measured twice, once before and once after some treatment or intervention, such as recording each patient's weight before and after a diet programme. For independent samples, testing H0 concerning mu1-mu2 proceeds through one of three parallel cases depending on what is known or assumed about the two population variances and how large the two samples are. When both populations are normal and their variances sigma1-squared and sigma2-squared are already known, the test statistic Z=[(X-bar-1-X-bar-2)-delta0]/sqrt(sigma1-squared/n1+sigma2-squared/n2) is exactly standard normal under H0, for any sample size whatsoever. When both samples are instead small (conventionally n1 and n2 both no larger than 30), the populations are normal, and their two variances, though unknown, are assumed EQUAL to some common value sigma-squared, the two separate sample variance estimates S1-hat-squared and S2-hat-squared must first be combined into a single pooled estimate, Sp-squared=[(n1-1)S1-hat-squared+(n2-1)S2-hat-squared]/(n1+n2-2), essentially a weighted average of the two sample variances weighted by their respective degrees of freedom; the resulting test statistic T=[(X-bar-1-X-bar-2)-delta0]/[Sp.sqrt(1/n1+1/n2)] then follows a t-distribution with v=n1+n2-2 degrees of freedom. When both samples are instead large (greater than 30), the Central Limit Theorem removes the need for either a normality assumption or an equal-variance assumption altogether, and the test statistic simply becomes Z=[(X-bar-1-X-bar-2)-delta0]/sqrt(S1-hat-squared/n1+S2-hat-squared/n2), using each sample's own separate variance estimate directly, with no pooling required. Dependent (paired) samples call for an entirely different, and in an important sense much SIMPLER, approach. Rather than working with the two original sets of measurements directly, the key manoeuvre is to collapse each matched pair down into a single number: the difference D=Y-X for that pair (for instance, weight-after minus weight-before, for each individual person in the study). Once this collapsing step has been carried out, the entire

two-sample comparison problem has been transformed into an ordinary ONE-sample problem, concerning only the population mean of these differences, denoted mu_D=mu2-mu1, and every formula and technique from the single-sample mean test developed earlier in this very chapter can then be applied directly and without modification, just substituted onto the sample of differences rather than onto the original raw measurements: specifically, the test statistic T=(D-bar-delta0)/(S_D-hat/sqrt(n)), calculated from the n paired differences' own sample mean D-bar and own sample standard deviation S_D-hat, follows a t-distribution with v=n-1 degrees of freedom, exactly mirroring the single-sample small-sample t-test structure. The genuine practical advantage of deliberately using a paired design, whenever it is actually feasible to do so, is that pairing allows each individual subject to effectively serve as their own personal control or baseline, which very efficiently strips out subject-to-subject variability that would otherwise remain baked into, and would inflate, the noise present in a comparison based on two entirely separate, unrelated independent samples. Consider, for example, comparing the effectiveness of two different weight-loss diets: if diet A were tested on one entirely separate group of people while diet B were tested on a completely different second group of people (an independent-samples design), any observed difference between the two groups' outcomes would be muddied and confounded together with all sorts of pre-existing individual differences between the particular people who happened to be randomly assigned into each of the two separate groups -- differences in starting weight, metabolism, activity level, genetics, and so on. If, instead, the very same group of individuals could somehow be tested on both diets sequentially (or, more practically, closely matched pairs of highly similar individuals could each be split one-to-a-diet), then each person's own individual baseline characteristics would automatically cancel out entirely when the paired difference D is calculated for that person, leaving only the genuine treatment effect of interest still remaining in the resulting differences -- and this systematic removal of extraneous subject-to-subject variability from the comparison is exactly why paired designs generally achieve substantially greater statistical power to correctly detect a real, genuinely existing difference, particularly whenever the available sample sizes are comparatively small, compared to what an equivalently-sized independent-samples design testing the very same underlying research question would typically be able to achieve.

Multiple Choice Questions (MCQs)

The null hypothesis H0 is best described as: (A) The claim the researcher hopes to prove (B) A statement of 'no effect' or the status quo, tested for possible rejection (C) Always a one-tailed statement (D) Only used for testing proportions

Correct answer: (B) A statement of 'no effect' or the status quo, tested for possible rejection. H0 is always a 'no effect'/status-quo statement, assumed true and tested for possible rejection.

A Type-I error occurs when: (A) H0 is accepted when it is false (B) H0 is rejected when it is actually true (C) H1 is rejected when it is true (D) The sample size is too small

Correct answer: (B) H0 is rejected when it is actually true. Type-I error = rejecting a true H0; its probability is alpha, the level of significance.

A Type-II error occurs when: (A) H0 is rejected when it is true (B) H0 is not rejected when H1 is actually true (C) The test statistic is miscalculated (D) Alpha is set too low

Correct answer: (B) H0 is not rejected when H1 is actually true. Type-II error = failing to reject H0 (accepting it) when H1 is actually true; its probability is beta.

If H1 is stated as theta < theta0, the appropriate test is: (A) Two-tailed (B) Right-tailed (C) Left-tailed (D) No test is possible

Correct answer: (C) Left-tailed. H1 with '<' indicates a definite decrease is being sought, requiring a one-sided LEFT-tailed test.

For a normal population with unknown variance and a small sample (n<=30), the correct test statistic for the mean is: (A) Z, using sigma (B) T, using S-hat, with v=n-1 degrees of freedom (C) Chi-square (D) F

Correct answer: (B) T, using S-hat, with v=n-1 degrees of freedom. Small sample, normal population, unknown variance: use the T statistic with n-1 degrees of freedom.

The continuity correction is applied when testing a hypothesis about: (A) The difference of two means (B) A population proportion, approximating the discrete binomial with the normal distribution (C) Paired samples (D) The pooled variance

Correct answer: (B) A population proportion, approximating the discrete binomial with the normal distribution. The continuity correction adjusts for approximating the discrete binomial (X or P) using the continuous normal distribution.

For comparing two independent means with small samples and equal but unknown variances, the correct test statistic is: (A) Z with pooled variance (B) T with pooled variance Sp^2, v=n1+n2-2 (C) Z with separate variances (D) T with v=n1+n2-1

Correct answer: (B) T with pooled variance Sp^2, v=n1+n2-2. This case requires the pooled variance estimate Sp^2 and a t-distribution with v=n1+n2-2 degrees of freedom.

Testing the difference between two means using paired (dependent) samples reduces to: (A) A two-sample Z-test (B) A one-sample t-test on the differences D=Y-X (C) A chi-square test (D) A test of two proportions

Correct answer: (B) A one-sample t-test on the differences D=Y-X. Pairing collapses the two-sample problem into an ordinary one-sample t-test performed on the differences.

To test H0: pi1=pi2 (equal proportions in two populations), the standard error is estimated using: (A) Separate P1 and P2 only (B) A pooled proportion pi-hat=(X1+X2)/(n1+n2) (C) The population variance (D) The t-distribution

Correct answer: (B) A pooled proportion pi-hat=(X1+X2)/(n1+n2). Testing pi1=pi2 specifically uses a pooled estimate of the single common proportion, pi-hat, combining both samples.

The level of significance, alpha, must be: (A) Calculated after the sample is collected (B) Fixed before the sample is drawn (C) Always equal to 0.5 (D) The same as the level of confidence

Correct answer: (B) Fixed before the sample is drawn. Alpha is chosen and fixed BEFORE sampling, so the decision rule cannot be influenced by the observed data.

Quick Revision Summary

- H0 = status quo/no-effect (tested for rejection) | H1 = research claim (accepted only if H0 rejected)

- Type-I error = reject true H0, probability alpha | Type-II error = accept false H0, probability beta
- Left-tailed: H1 uses '<' | Right-tailed: H1 uses '>' | Two-tailed: H1 uses '!='
- 7 steps: state hypotheses -> set alpha -> pick test statistic -> find critical value(s) -> rejection region -> decision rule -> compute & conclude
- Mean test: sigma known -> Z; small n normal sigma unknown -> T (v=n-1); large n -> Z (CLT)
- Proportion test: Z=(P-pi0)/sqrt[pi0(1-pi0)/n]; continuity correction adds/subtracts 0.5
- Diff of means, independent: known var -> Z; small n equal var -> pooled T (v=n1+n2-2); large n -> Z
- Diff of means, paired/dependent: reduces to 1-sample T-test on differences D=Y-X, v=n-1
- Diff of proportions: general delta0 -> separate P1,P2 in SE; testing pi1=pi2 -> pooled pi-hat

Exam Tips

- Always write H0 with the equality sign (>=, <=, or =) and H1 as the strict, direction-matching complement
- Match the tail of the test to H1's inequality: '<' = left tail, '>' = right tail, '!=' = two tails (split alpha/2 each side)
- For mean tests, check THREE things before picking Z or T: is sigma known, is the population normal, is n large or small
- For paired-sample problems, compute the differences FIRST (consistently: always after-minus-before, or always A-minus-B) before doing anything else
- When testing pi1=pi2 specifically, don't forget to pool: use pi-hat in the standard error, not separate p1 and p2
- State your final conclusion in plain managerial language, not just 'reject H0' -- explain what that means for the actual question being asked