Chapter 10

Statistical Inference for Two Samples

Chapter 10: Statistical Inference for Two Samples

LEARNING OBJECTIVES

10.1 Inference on the Difference in Means of Two Normal Distributions, Variance Known

10.2 Inference on the Difference in Means of Two Normal Distributions, Variance Unknown

10.6 Inference on the Two Proportions

10.1. Inference on the Difference in Means of Two Normal Distributions, Variance Known

Assumptions for Two-Sample Inference

(1) $X_{11}, X_{12}, \ldots, X_{1n1}$ is a random sample from population 1.
(2) $X_{21}, X_{22}, \ldots, X_{2n2}$ is a random sample from population 2.
(3) The two populations represented by X_1 and X_2 are independent.
(4) Both populations are normal.

Tests on the Difference in Means, Var(X) known

Null hypothesis: H_0: $\mu_1 - \mu_2 = \Delta_0$

Test statistic: $$Z_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Alternative hypothesis	P-Value	Rejection criteria for Fixed-Level Tests
H_1: $\mu_1 - \mu_2 \neq \Delta_0$	$P = 2[1 - \Phi(\lvert z_0\rvert)]$	$\lvert z_0\rvert > z_{\alpha/2}$
H_1: $\mu_1 - \mu_2 > \Delta_0$	$P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
H_1: $\mu_1 - \mu_2 < \Delta_0$	$P = \Phi(z_0)$	$z_0 < -z_\alpha$

σ^2 known

Example 2

A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2; the 20 specimens are painted in random order. The two sample average drying times are $\overline{x}_1 = 121$ minutes and $\overline{x}_2 = 112$ minutes, respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

Probability of type II error β for a two-sided test

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

Sample Siz for a two-sided test

For the two-sided alternative hypothesis with significance level α, the sample size $n_1 = n_2 = n$ required to detect a true difference in means of Δ with power at least 1 - β is

$$n \simeq \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2 \left(\sigma_1^2 + \sigma_2^2\right)}{\left(\Delta - \Delta_0\right)^2}$$

Sample Siz for a One-sided test

For a one-sided alternative hypothesis with significance level α, the sample size $n_1 = n_2 = n$ required to detect a true difference in means of Δ(≠ Δ0) with power at least 1 - β is

$$n \simeq \frac{\left(z_{\alpha} + z_{\beta}\right)^2 \left(\sigma_1^2 + \sigma_2^2\right)}{\left(\Delta - \Delta_0\right)^2}$$

Confidence Interval on the Difference in Means, Variances Known

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2 from two independent normal populations with known variances σ_1^2 and σ_2^2, respectively, a $100(1-\alpha)\%$ **confidence interval for $\mu_1 - \mu_2$** is

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10.7)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

Sample Size for a Confidence Interval on the Difference in Means, Variances Known

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (\sigma_1^2 + \sigma_2^2) \quad (10.8)$$

One-Sided Upper-Confidence Bound

$$\mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{10.9}$$

One-Sided Lower-Confidence Bound

$$\bar{x}_1 - \bar{x}_2 - z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \tag{10.10}$$

Exercise

Consider the hypothesis test H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 \neq \mu_2$ with known variances $\sigma_1 = 10$ and $\sigma_2 = 5$. Suppose that sample sizes $n_1 = 10$ and $n_2 = 15$ and that $\overline{x}_1 = 4.7$ and $\overline{x}_2 = 7.8$. Use $\alpha = 0.05$.

(a) Test the hypothesis and find the *P*-value.

(b) Explain how the test could be conducted with a confidence interval.

(c) What is the power of the test in part (a) for a true difference in means of 3?

(d) Assume that sample sizes are equal. What sample size should be used to obtain $\beta = 0.05$ if the true difference in means is 3? Assume that $\alpha = 0.05$

10.2 Inference on the Difference in Means of Two Normal Distributions, Variance Unknown

Tests on the Difference in Means, Var(X) Unknown

Case 1: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Pooled Estimator of Variance

The **pooled estimator** of σ^2, denoted by S_p^2, is defined by

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Tests on the Difference in Means, Var(X) Unknown

Null hypothesis: H_0: $\mu_1 - \mu_2 = \Delta_0$

Test statistic: $$T_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Alternative hypothesis	P-Value	Rejection criteria for Fixed-Level Tests
H_1: $\mu_1 - \mu_2 \neq \Delta_0$	$P = 2[1 - \Phi(\lvert z_0\rvert)]$	$\lvert t_0\rvert > t_{\alpha/2;\, n1+n2-2}$
H_1: $\mu_1 - \mu_2 > \Delta_0$	$P = 1 - \Phi(z_0)$	$t_0 > t_{\alpha;\, n1+n2-2}$
H_1: $\mu_1 - \mu_2 < \Delta_0$	$P = \Phi(z_0)$	$t_0 < - t_{\alpha;\, n1+n2-2}$

Example

Two catalysts are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, catalyst 1 is currently used; but catalyst 2 is acceptable. Because catalyst 2 is cheaper, it should be adopted, if it does not change the process yield. A test is run in the pilot plant and results in the data shown in Table 10.1. Figure 10.2 presents a normal probability plot and a comparative box plot of the data from the two samples. Is there any difference in the mean yields? Use $\alpha = 0.05$, and assume equal variances.

Observation Number	1	2	3	4	5	6	7	8
Catalyst 1	91.50	94.18	92.18	95.39	91.79	89.07	94.72	89.21
Catalyst 2	89.19	90.95	90.46	93.21	97.19	97.04	91.07	92.75

Example

An experiment was performed to compare the abrasive wear of two different laminated materials. Twelve pieces of material 1 were tested by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 units with a sample standard deviation of 4, while the samples of material 2 gave an average of 81 with a sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units? Assume the populations to be approximately normal with equal variances.

Case 2: $\sigma_1^2 \neq \sigma_2^2$

If H_0: $\mu_1 - \mu_2 = \Delta_0$ is true, the statistic

$$T_0^* = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

is distributed approximately as t with degrees of freedom given by

$$v = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}}$$

If v is not an integer, round down to the nearest integer

Example

potential health risk. An article in the *Arizona Republic* (May 27, 2001) reported drinking water arsenic concentrations in parts per billion (ppb) for 10 metropolitan Phoenix communities and 10 communities in rural Arizona. The data follow:

Metro Phoenix	**Rural Arizona**
$(\bar{x}_1 = 12.5,\ s_1 = 7.63)$	$(\bar{x}_2 = 27.5,\ s_2 = 15.3)$
Phoenix, 3	Rimrock, 48
Chandler, 7	Goodyear, 44
Gilbert, 25	New River, 40
Glendale, 10	Apache Junction, 38
Mesa, 15	Buckeye, 33
Paradise Valley, 6	Nogales, 21
Peoria, 12	Black Canyon City, 20
Scottsdale, 25	Sedona, 12
Tempe, 15	Payson, 1
Sun City, 7	Casa Grande, 18

We wish to determine whether any difference exists in mean arsenic concentrations for metropolitan Phoenix communities and for communities in rural Arizona.

Confidence Interval on the Difference in Means, Variances Unknown

Case 1: Confidence Interval on the Difference in Means, Variances Unknown and Equal

If $\bar{x}_1, \bar{x}_2, s_1^2$, and s_2^2 are the sample means and variances of two random samples of sizes n_1 and n_2, respectively, from two independent normal populations with unknown but equal variances, a $100(1-\alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{\alpha/2, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{10.19}$$

where $s_p = \sqrt{[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)}$ is the pooled estimate of the common population standard deviation, and $t_{\alpha/2, n_1+n_2-2}$ is the upper $\alpha/2$ percentage point of the t distribution with $n_1 + n_2 - 2$ degrees of freedom.

Confidence Interval on the Difference in Means, Variances Unknown

Case 2: Approximate Confidence Interval on the Difference in Means, Variances Unknown and Not Assumed Equal

If $\bar{x}_1, \bar{x}_2, s_1^2$, and s_2^2 are the means and variances of two random samples of sizes n_1 and n_2, respectively, from two independent normal populations with unknown and unequal variances, an approximate $100(1-\alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (10.20)$$

where v is given by Equation 10.16 and $t_{\alpha/2,v}$ is the upper $\alpha/2$ percentage point of the t distribution with v degrees of freedom.

EXAMPLE.

An article in the journal *Hazardous Waste and Hazardous Materials* (1989, Vol. 6) reported the results of an analysis of the weight of calcium in standard cement and cement doped with lead. Reduced levels of calcium would indicate that the hydration mechanism in the cement is blocked and would allow water to attack various locations in the cement structure. Ten samples of standard cement had an average weight percent calcium of $\bar{x}_1 = 90$ with a sample standard deviation of $s_1 = 5.0$, and 15 samples of the lead-doped cement had an average weight percent calcium of $\bar{x}_2 = 87$ with a sample standard deviation of $s_2 = 4.0$.

Assume that weight percent calcium is normally distributed and find a 95% confidence interval on the difference in means, $\mu_1 - \mu_2$, for the two types of cement. Furthermore, we assume that both normal populations have the same standard deviation.

10.6. Inference on Two Population Proportions

Test Statistic for the Difference of Two Population Proportions

Null hypothesis: H_0: $p_1 = p_2$

Test statistic: $$Z_0 = \frac{P_1 - P_2}{\sqrt{P\left(1-P\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Alternative hypothesis	P-Value	Rejection criteria
H_1: $p_1 \neq p_2$	$P = 2[1 - \Phi(\lvert z_0\rvert)]$	$\lvert z_0\rvert > z_{\alpha/2}$
H_1: $p_1 > p_2$	$P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
H_1: $p_1 < p_2$	$P = \Phi(z_0)$	$z_0 < - z_\alpha$

Example

Extracts of St. John's Wort are widely used to treat depression. An article in the April 18, 2001, issue of the *Journal of the American Medical Association* ("Effectiveness of St. John's Wort on Major Depression: A Randomized Controlled Trial") compared the efficacy of a standard extract of St. John's Wort with a placebo in 200 outpatients diagnosed with major depression. Patients were randomly assigned to two groups; one group received the St. John's Wort, and the other received the placebo. After 8 weeks, 19 of the placebo-treated patients showed improvement, and 27 of those treated with St. John's Wort improved. Is there any reason to believe that St. John's Wort is effective in treating major depression? Use $\alpha = 0.05$.

Example

A vote is to be taken among the residents of a town and the surrounding county to determine whether a proposed chemical plant should be constructed. The construction site is within the town limits, and for this reason many voters in the county believe that the proposal will pass because of the large proportion of town voters who favor the construction. To determine if there is a significant difference in the proportions of town voters and county voters favoring the proposal, a poll is taken. If 120 of 200 town voters favor the proposal and 240 of 500 county residents favor it, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use an $\alpha = 0.05$ level of significance.

Type II Error and Choice of Sample Size

Two-sided test

$$\beta = \Phi\left[\frac{z_{\alpha/2}\sqrt{\overline{pq}\left(1/n_1+1/n_2\right)}-\left(p_1-p_2\right)}{\sigma_{p_1-p_2}}\right]$$

$$-\Phi\left[\frac{-z_{\alpha/2}\sqrt{\overline{pq}\left(1/n_1+1/n_2\right)}-\left(p_1-p_2\right)}{\sigma_{p_1-p_2}}\right]$$

where

$$\bar{p}=\frac{n_1p_1+n_2p_2}{n_1+n_2},\ \bar{q}=\frac{n_1\left(1-p_1\right)+n_2\left(1-p_2\right)}{n_1+n_2}$$

and

$$\sigma_{p_1-p_2}=\sqrt{\frac{p_1\left(1-p_1\right)}{n_1}+\frac{p_2\left(1-p_2\right)}{n_2}}$$

Type II Error and Choice of Sample Size

One-sided test: $p_1 > p_2$

$$\beta = \Phi\left[\frac{z_\alpha\sqrt{\overline{pq}\left(1/n_1+1/n_2\right)}-\left(p_1-p_2\right)}{\sigma_{p_1-p_2}}\right]$$

One-sided test: $p < p_0$

$$\beta = 1-\Phi\left[\frac{-z_\alpha\sqrt{\overline{pq}\left(1/n_1+1/n_2\right)}-\left(p_1-p_2\right)}{\sigma_{p_1-p_2}}\right]$$

Sample Size Formulas

Two-sided test

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

where $q_1=1-p_1$ and $q_2 = 1-p_2$.

Example

Two different types of injection-molding machines are used to form plastic parts. A part is considered defective if it has excessive shrinkage or is discolored. Two random samples, each of size 300, are selected, and 15 defective parts are found in the sample from machine 1, and 8 defective parts are found in the sample from machine 2.

(a) Is it reasonable to conclude that both machines produce the same fraction of defective parts, using $\alpha = .0\ 05$? Find the *P*-value for this test.

(b) Suppose that $p_1 = 0.05$ and $p_2 = 0.01$. With the sample sizes given here, what is the power of the test for this two-sided alternate?

(c) Suppose that $p_1 = 0.05$ and $p_2 = 0.01$. Determine the sample size needed to detect this difference with a probability of at least 0.9.

Confidence Interval on the Difference in Population Proportions

Approximate Confidence Interval on the Difference in Population Proportions

If $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions of observations in two independent random samples of sizes n_1 and n_2 that belong to a class of interest, an approximate two-sided $100(1-\alpha)\%$ confidence interval on the difference in the true proportions $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \le \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \quad (10.41)$$

where $Z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.