Probability and Statistics for Engineers

Clémonell Bilayi-Biakana, PhD.
Departement of Mathematics and Statistics
University of Ottawa

July 22, 2021.

Outline

Hypothesis Testing Problem

Motivating Example

1. *In a courtroom, a defendant standing trial for a crime is assumed innocent until proven guilty.*
2. *It is the job of the prosecution to present evidence showing that the defendant is guilty beyond a reasonable doubt. It is the job of the defense to challenge this evidence to establish a reasonable doubt.*
3. *The jury weighs the evidence and makes a decision.*

Hypothesis Testing Problem

This example can be decomposed into 3 major parts:

1. Formulation of hypotheses.

$$\begin{cases} H_0 : & \text{The defendant is innocent.} \\ H_1 : & \text{The defendant is guilty.} \end{cases}$$

2. Fact-checking-evidence.
   In a courtroom, a defendant is assumed to be innocent until proven guilty. It is the job of the prosecution to present evidence showing that the defendant is guilty beyond a reasonable doubt. It is the job of the defense to challenge this evidence to establish a reasonable doubt.
3. Decision making.
   The jury weighs the evidence and makes a decision.

Hypothesis Testing Problem

- Let X be a random variable such that $X \sim f_\theta(x)$, where $\theta \in \Theta$ is an unknown parameter. The sample parameter $\Theta = \Theta_0 \dot{\cup} \Theta_1 \subseteq \mathbb{R}$. Note that $\Theta_1 = \Theta_0^c$. We assume that the functional form of the pdf or the pmf of X is known but unknown for θ.
- Researchers often have preconceived ideas about what the values of these parameters might be and wish to test whether the data conform to these ideas.
- In statistical inference, a statement about the parameters of one or more populations is called **statitiscal hypothesis**.
- The decision-making procedure about the hypothesis is called **hypothesis testing**.

Hypothesis Testing Problem

In many practical problems an experimenter is interested in testing the validity of an assertion about the unknown parameter θ. To proceed, an experimenter formulates the statistical hypotheses as follows:

$$(H) : \begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1. \end{cases} \tag{1}$$

- $H_0 : \theta \in \Theta_0$ is called **null hypothesis**
- $H_1 : \theta \in \Theta_1$ is called **alternative hypothesis**.

The null hypothesis corresponds to the statement of "no difference" whereas the alternative represents "change".

Hypothesis Testing Problem

Some Facts

- *If* Θ_0 *is* **simple**, *that is* Θ_0 *consists of a single point* θ_0, *then* $H_0 : \theta \in \Theta_0$ *is called* **simple hypothesis**.
- *The probability distribution of* X *is completely specified under a simple hypothesis.*
- *If* Θ_0 *is* **composite**, *that is* Θ_0 *consists of more than a single point* θ_0, *then* $H_0 : \theta \in \Theta_0$ *is called* **composite hypothesis**. *Usually the parameter of interest* θ *has a range of values.*
- *These two definitions are applicable to* $H_1 : \theta \in \Theta_1$ *depending on whether or not* Θ_1 *is simple.*

Hypothesis Testing

Let $X \sim f(x, \theta)$, where θ is unkown parameter and θ_0, a fixed value of θ. The following are the different types of hypothesis tests that we will deal with in this course:

1. a **two-tailed test:** $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.
2. a **right-tailed test:** $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.
3. a **left-tailed test:** $H_0 : \theta = \theta_0$ versus $H_1 : \theta < \theta_0$.

A two tailed-test is also known as a **two-sided test**. Left-tailed and right-tailed tests fall into the class of **one-sided tests**.

Hypothesis Testing Problem

Formulation (Hypothesis Testing Problem)

Let X be a random variable such that $X \sim f_\theta(x)$, where $\theta \in \Theta \subset \mathbb{R}$ is an unknown population parameter. Consider performing the following hypothesis test:

$$(H) : \begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1, \end{cases} \tag{2}$$

where Θ_0 is a subset of Θ and $\Theta_1 = \Theta_0^c$.

- *Given the sample point $\boldsymbol{x} = (x_1, x_2, ..., x_n)$,*
- *find a **decision rule** that will lead to a decision to reject or fail to reject the null hypothesis. The sample quantity $T(\boldsymbol{x})$ on which the decision to support H_0 or H_1 is based is called **test statistic**.*

Hypothesis Testing

Steps in Performing a Hypothesis Test

1. Formulate the **statistical hypotheses**: H_0 and H_1. A statistical hypothesis is a conjecture about a population parameter. This conjecture may or may not be true.
2. Compute the **test statistic** and identify its sampling distribution. This should be done under the assumption that the null hypothesis H_0 is true.
3. Make a decision. This is done either by computing either the **critical value** or the p-**value**, that is the likelihood of obtaining a test statistic in the direction of the alternative hypothesis H_1 when the null hypothesis H_0 is true. Like in interval estimation, hypothesis testing is done with a given significance level α.

Hypothesis Testing

In the courtroom, the jury can possibly make the following decisions.

i) The jury concludes that there is enough evidence to convict the defendant. This “guilty verdict” is synonymous with rejecting H_0.

ii) The jury concludes that there is not enough evidence to conclude beyond a reasonable doubt that the person is guilty. Notice that they do not conclude that the person is innocent. This not “guilty verdict” is synonymous with failing to reject H_0.

Hypothesis Testing

These two verdicts are similar to the two conclusions that are possible in hypothesis testing:

1. **Reject the null hypothesis** H_0, that is there is sufficient evidence to reject the claim made in H_0.
2. **Fail to reject the null hypothesis** H_0, that is there is not sufficient evidence to reject the claim made in H_0.

iii) The jury convicts an innocent defendant.
In hypothesis testing, such a decision is referred to as **type I error**. A type I error occurs if one rejects the null hypothesis when it is true. The probability of making a type I error is:

$$\alpha = P(\text{reject } H_0 \| H_0 \text{ is true}).$$

iv) The jury says a defendant is not guilty when he or she really is.
In hypothesis testing, such a decision is referred to as **type II error**. A type II error occurs if one does not reject the null hypothesis when it is false. The probability of making a type II error is:

$$\beta = P(\text{fail to reject } H_0 \| H_0 \text{ is false}).$$

The possible decisions and errors that occur when performing a hypothesis testing are summarized in the next table.

Hypothesis Testing Problem

The possible decisions and errors made by an experimenter when performing hypothesis test (1) are displayed in the table:

	H_0 true	H_1 true
Reject H_0	Type I Error	Correct Decision
Do Not Reject H_0	Correct Decision	Type II Error

Definitions

- *A* **type I error** *or* **error of first kind** *occurs if an experimenter rejects $H_0 : \theta \in \Theta_0$ when it is true.*
- *A* **type II error** *or* **error of second kind** *occurs if an experimenter fails to reject $H_0 : \theta \in \Theta_0$ when it is false.*

Hypothesis Testing: One-Sample Inference
Hypothesis Testing for a Proportion

- Let p be the proportion of a population with a certain characteristic X.
- Let $(X_1, X_2, \ldots, X_n)$, $(n > 30)$ denote a large sample, Y denote the number of individuals with characteristic X and $\hat{P}_n = Y/n$, the sample proportion.
- Assume that $np > 5$ and $n(1-p) > 5$.

Consider performing one of the following hypothesis tests:

1. $H_0 : p = p_0$ versus $H_1 : p \neq p_0$
2. $H_0 : p = p_0$ versus $H_1 : p > p_0$
3. $H_0 : p = p_0$ versus $H_1 : p < p_0$.

Hypothesis Testing: One-Sample Inference
Hypothesis Testing for a Proportion

with a significance level of α. Under H_0, the test statistic

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{p_0(1 - p_0)/n}} \approx \mathcal{N}(0, 1).$$

Let $z_0 = \hat{p} - p_0/\sqrt{p_0(1 - p_0)/n}$ denote the observed value of Z_0.

1. If $H_0 : p = p_0$ versus $H_1 : p \neq p_0$, then reject H_0 if $|z_0| > z_{\alpha/2}$ or p-value$= 2P\left(Z_0 > |z_0|\right) < \alpha$.
2. If $H_0 : p = p_0$ versus $H_1 : p > p_0$, then reject H_0 if $z_0 > z_\alpha$ or p-value$= P\left(Z_0 > z_0\right) < \alpha$.
3. If $H_0 : p = p_0$ versus $H_1 : p < p_0$, then reject H_0 if $z_0 < -z_\alpha$ or p-value$= P\left(Z_0 < z_0\right) < \alpha$.

Hypothesis Testing: One-Sample Inference

Hypothesis Testing for a Proportion

Exercise

A manufacturer claims that a new beetle trap attracts and kills more than 90% of the beetles that come close to it. We would like to test this claim. A sample of 900 beetles is available from which 825 were attracted to the trap. Is there enough evidence that supports the claim? (Use the significance level $\alpha = 0.05$.)

Solution

Let p denote the proportion of beetles attracted and killed by the new trap. We want to test the following hypotheses:

$$H_0 : p = 0.9 \text{ versus } H_1 : p > 0.9,$$

with significance level $\alpha = 0.05$. *The point estimate of p is* $\hat{p} = 825/900 = 0.917$. *Under* H_0, *the test statistic*

$$Z_0 = \frac{\hat{P} - 0.9}{\sqrt{(0.9)(0.1)/900}} \approx \mathcal{N}(0, 1).$$

Solution (Cont'd)

The value of the test statistic for this sample is

$$z_0 = \frac{\hat{p} - 0.9}{\sqrt{(0.9)(0.1)/900}} = \frac{0.917 - 0.9}{\sqrt{(0.9)(0.1)/900}} = 1.7$$

Therefore using the standard normal, we get

$$p - value = P(Z_0 \geq 1.7) = 0.0446$$

Since the p-value is smaller than 0.05*, we reject* H_0*. The data does not support the claim made in the null hypothesis.*

Hypothesis Testing: One-Sample Inference

Hypothesis Testing for the Mean

Let $(X_1, X_2, \ldots, X_n)$ be a sample from $X \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Consider performing one of the following tests:

1. $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$
2. $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$
3. $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$.

with a significance level of α. Under H_0, the test statistic

$$Z_0 = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Hypothesis Testing: One-Sample Inference

Hypothesis Testing for the Mean

Let $z_0 = \sqrt{n}(\bar{x} - \mu_0)/\sigma$ denote the observed value of Z_0.

1. If $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, then reject H_0 if $|z_0| > z_{\alpha/2}$ or p-value$= 2P\left(Z_0 > |z_0|\right) < \alpha$.
2. If $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, then reject H_0 if $z_0 > z_\alpha$ or p-value$= P\left(Z_0 > z_0\right) < \alpha$.
3. If $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$, then reject H_0 if $z_0 < -z_\alpha$ or p-value$= P\left(Z_0 < z_0\right) < \alpha$.

Hypothesis Testing: One-Sample Inference

Hypothesis Testing for the Mean: Large Samples

Let $(X_1, X_2, \dots, X_n)$, $(n > 30)$, be a large random sample from a population whose parameter of interest is μ. Consider performing one of the following hypothesis tests:

1. $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$
2. $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$
3. $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$.

with a significance level of α. Under H_0, the test statistic

$$Z_0 = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \approx \mathcal{N}(0, 1).$$

Hypothesis Testing: One-Sample Inference

Hypothesis Testing for the Mean: Large Samples

Let $z_0 = \sqrt{n}(\bar{x} - \mu_0)/\sigma$ denote the observed value of Z_0.

1. If $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, then reject H_0 if $|z_0| > z_{\alpha/2}$ or p-value$= 2P\left(Z_0 > |z_0|\right) < \alpha$.
2. If $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, then reject H_0 if $z_0 > z_\alpha$ or p-value$= P\left(Z_0 > z_0\right) < \alpha$.
3. If $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$, then reject H_0 if $z_0 < -z_\alpha$ or p-value$= P\left(Z_0 < z_0\right) < \alpha$.

Exercise

The medical rehabilitation education foundation reports that the average cost of rehabilitation for stroke victims is \$ 24,672. To see if the average cost of rehabilitation is different at a particular hospital, a researcher selects a random sample of 35 stroke victims at the hospital and finds that the average cost of their rehabilitation is \$ 25, 226. The standard deviation of the population is \$ 3251. At $\alpha = 0.01$*, can it be concluded that the average cost of the stroke rehabilitation at a particular hospital is different from \$24,672?*

Solution

Let μ denote the average cost of the stroke rehabilitation at a particular hospital. This statistical problem consists of testing: $H_0 : \mu = 24,672$ versus $H_1 : \mu \neq 24,672$, with a significance level of $\alpha = 0.01$. Under H_0, the test statistic

$$Z_0 = \frac{\bar{X} - 24,672}{3251/\sqrt{35}} \approx \mathcal{N}(0,1).$$

Therefore, the observed value of Z_0, is

$$z_0 = \frac{\sqrt{35}(25,226 - 24,672)}{3251} \approx 1.01$$

Solution (Cont'd)

Since $\alpha = 0.01$, *then* $\alpha/2 = 0.005$. *Therefore,*

$$P(Z_0 > z_{0.005}) = 0.05 \Rightarrow z_{0.005} = 2.57.$$

It turns out $|z_0| < z_{\alpha/2}$, *that is* $1.01 < 2.57$. *Hence, we do not reject* H_0. *This means that there is not enough evidence to support the claim made by the medical rehabilitation education foundation that: "the average cost of rehabilitation for stroke victims is \$ 24,672".*

Remark

Notice that the same conclusion is reachable using the p-value approach. In fact since $Z_0 \sim \mathcal{N}(0,1)$, then

$$\begin{aligned} p - value &= 2P(Z_0 > |z_0|) \\ &= 2P(Z_0 > 1.01) \\ &= 2[1 - \Phi(1.01)] \\ &= 2(1 - 0.8438) \\ &= 0.3124\ . \end{aligned}$$

Since the p-value > 0.01, then we do not reject H_0.

Hypothesis Testing: One-Sample Inference

Hypothesis testing for the Mean: Small Samples

Let $(X_1, X_2, \ldots, X_n)$, $(n \leq 30)$, be a sample from $X \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is unknown. Consider performing one of the following hypothesis tests:

1. $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$
2. $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$
3. $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$.

with a significance level of α. Under H_0, the test statistic

$$T_0 = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \sim T_{n-1}.$$

Hypothesis Testing: One-Sample Inference

Hypothesis testing for the Mean: Small Samples

Let $t_0 = \sqrt{n}(\bar{x} - \mu_0)/s$ denote the observed value of T_0.

- If $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, then reject H_0 if $|t_0| > t_{n-1;\alpha/2}$ or p-value$= 2P\left(T_0 > |t_0|\right) < \alpha$.
- If $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, then reject H_0 if $t_0 > t_{n-1;\alpha}$ or p-value$= P\left(T_0 > t_0\right) < \alpha$.
- If $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$, then reject H_0 if $t_0 < -t_{n-1;\alpha}$ or p-value$= P\left(T_0 < t_0\right) < \alpha$.

Exercise

The systolic blood pressure level in a certain hypertensive population is approximately equal to the value 130 *mm Hg. A new drug is developed to reduce the systolic blood pressure levels in this population under the value* 130*. Suppose that the new drug has been administered to a random sample of* 16 *patients. The observed values for this sample yielded a sample mean* $\bar{x} = 123.7$ *and a sample standard deviation* $s = 15.4$*. Is there enough evidence that the drug is efficient in reducing the systolic blood pressure? Assume that the population distribution is normal and use the significance level* $\alpha = 0.01$*.*

Solution

Let μ denote the mean level of systolic blood pressure for the patients who were administered the drug. The statistical problem described in assumptions is equivalent to performing the following hypothesis test: $H_0 : \mu = 130$ vs $H_1 : \mu < 130$, with a significance level of $\alpha = 0.01$. Under H_0,

$$T_0 = \frac{\bar{X} - 130}{S/\sqrt{16}} \sim T_{15}.$$

The observed value of the test statistic is

$$t_0 = \frac{\bar{x} - 130}{s/\sqrt{16}} = \frac{123.7 - 130}{15.4/\sqrt{16}} = -1.636$$

Hence, p-value $= P(T_0 < -1.636) = P(T_0 > 1.636)$.

Solution (Cont'd)

Looking across the row corresponding to 15 degrees of freedom in the t-table, we obtain that

$$1.341 < 1.636 < 1.753$$
$$P(T > 1.753) < P(T > 1.636) < P(T > 1.341)$$
$$0.05 < P(T > 1.636) < 0.10$$
$$0.05 < p - value < 0.10$$

Therefore, the p-value > 0.01. *Hence, we fail to reject* H_0. *We conclude that there is not enough evidence that the new drug is efficient in reducing the systolic blood pressure.*

Exercise

Measurements of blood viscosity were made on laboratory mice. A normal value should be close to 3.95*. Researchers which are testing a new drug suspect that this could have modified their blood viscosity level. Levels which are either too small or too large are not acceptable. A sample of* 9 *mice yielded* $\bar{x} = 4.25$ *and* $s = 0.6$*. Is there enough evidence that the average level of viscosity is different than 3.95? Assume that the level of viscosity is normally distributed and use the level* $\alpha = 0.05$*.*

Solution

Let μ be the average viscosity level. This statistical problem is equivalent to testing: $H_0 : \mu = 3.95$ versus $H_1 : \mu \neq 3.95$, with the significance level $\alpha = 0.05$. Under H_0, the test statistic

$$T_0 = \frac{\bar{X} - 3.95}{S/3} \sim T_8.$$

The observed value of the test statistic is

$$t_0 = \frac{\bar{x} - 3.95}{s/3} = \frac{4.25 - 3.95}{0.6/3} = 1.5$$

Solution (Cont'd)

Hence, p-value $= 2P(T_0 > 1.5)$. *Looking across the row corresponding to 8 degrees of freedom in t-table, we observe that* $1.397 < 1.5 < 1.860$. *Therefore,*

$$P(T > 1.860) < P(T > 1.5) < P(T > 1.397)$$
$$2(0.05) < 2P(T > 1.5) < 2(0.10)$$
$$0.10 < p - value < 0.20$$

Since p-value > 0.05, *we cannot reject* H_0. *There is not enough evidence that* μ *is different that 3.95.*

Exercise

The standard permissible exposure to benzene in the oil refining industry is 1 part per million (ppm). An average value larger that 1 ppm is felt to be dangerous. An industrial hygienist at a specific oil company measured the benzene exposure levels of 25 *workers. He obtained a sample mean* $\bar{x} = 1.03$ *ppm with a sample standard deviation* $s = 0.075$ *ppm. Using this data, is there enough evidence that the level of exposure is larger than 1 ppm? Justify your answer based on the p-value and use the significance level* $\alpha = 0.05$*. The level of exposure is assumed to be normally distributed.*

Solution

Let μ be the average level of exposure. The statistical problem described in assumptions can be formulated as follows: $H_0 : \mu = 1$ versus $H_1 : \mu > 1$, with a significance level of $\alpha = 0.05$. Under H_0, the test statistic

$$T_0 = \frac{5(\bar{X} - 1)}{S} \sim T_{24}$$

The observed value of this test statistic is

$$t_0 = \frac{\bar{x} - 1}{s/5} = \frac{1.03 - 1}{0.075/5} = 2.00$$

Hence, $p - value = P(T_0 > 2.00)$.

Solution

Using the t-table, we get $P(T > 1.711) = 0.05$ *and* $P(T > 2.064) = 0.025$. *We can conclude that*

$$0.025 < p - value < 0.05.$$

Since p-value < 0.05, *we reject* H_0. *There is enough evidence to reject the claim made in* H_0.

Hypothesis Testing: Two-Sample Inference

Normal Populations with Equal Variances

Let $(X_1, X_2, \ldots, X_n)$ be a sample from $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. Let $(Y_1, Y_2, \ldots, Y_m)$ be a sample from $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Further, we assume that $X \perp\!\!\!\perp Y$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$, with σ^2- **unknown**. The sample sizes are assumed to be small. Consider performing one of the following hypothesis tests:

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2$$
$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 > \mu_2$$
$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 < \mu_2,$$

with a significance level α. Under H_0, the test statistic

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2(1/n + 1/m)}} \sim T_{n+m-2},$$

Hypothesis Testing: Two-Sample Inference
Normal Populations with Equal Variances

where the **pooled sample variance** estimate is:

$$s_p^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$$

Let t_0 denote the observed value of T_0.

- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, then reject H_0 if $|t_0| > t_{n+m-2;\alpha/2}$ or p-value$= 2P\left(T_0 > |t_0|\right) < \alpha$.
- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$, then reject H_0 if $t_0 > t_{n+m-2;\alpha}$ or p-value$= P\left(T_0 > t_0\right) < \alpha$.
- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$, then reject H_0 if $t_0 < -t_{n+m-2;\alpha}$ or p-value$= P\left(T_0 < t_0\right) < \alpha$.

Exercise

We want to compare the average scores μ_1 and μ_2 on a standardized test in mathematics taken by students from large and small high schools. The experiment yielded the following information:

large high schools	*small high schools*
$n_1 = 9$	$n_2 = 15$
$\bar{x}_1 = 81.31$	$\bar{x}_2 = 78.61$
$s_1^2 = 60.76$	$s_2^2 = 48.24$

Exercise (Cont'd)

1. *Compute a 95% confidence interval for the average difference between the scores of large and small high schools. Using this interval, can we say that there is enough evidence that the average score is higher in a large school than in a small school?*
2. *Compare the means μ_1 and μ_2 from the point of view of hypothesis testing. Is there enough evidence that μ_1 is larger than μ_2?*

Assume that the populations are normal with equal variances.

Solution

The pooled variance for the two samples is:

$$s_p^2 = \frac{8(60.76) + 14(48.24)}{9 + 15 - 2} = 52.79$$

From the t-table, we see that the value t such that $P(-t \leq T \leq t) = 0.95$ is $t = 2.074$, where T has a student T distribution with 22 degrees of freedom.

1) A 95% confidence interval for $\mu_1 - \mu_2$ is:

$$I_{\mu_1 - \mu_2} = \left[81.31 - 78.61 \pm 2.074 \sqrt{(52.79)\left(\frac{1}{9} + \frac{1}{15}\right)} \right]$$
$$= [-3.65;\ 9.05].$$

Since the interval contains 0 we cannot conclude that there

Solution (Cont'd)

is a difference between the two scores.
2) We would like to test $H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 > \mu_2$*, with significance level* 0.05*. Under* H_0*, the test statistic*

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{(1/9 + 1/15)}} \sim T_{22},$$

The observed value of the test statistic is

$$t_0 = \frac{81.31 - 78.61}{\sqrt{(52.79)\left(\frac{1}{9} + \frac{1}{15}\right)}} = 0.881$$

Since $0.05 < 0.15 < p - value = P(T > 0.881) < 0.20$*, then we fail to reject the null hypothesis* H_0*.*

Hypothesis Testing: Two-Sample Inference

Normal Populations with Equal Variances

Exercise

It is believed that nutritional deprivation affects various components of the immune system, such as the tuberculin skin reactivity. In a study, a sample of 8 Sprague-Dawley male rats were fed with a normal diet consisting of 18% protein. A state of malnutrition was induced in another sample of 8 rats, which were fed with a diet consisting of only 5% protein. After 4 weeks, the rats were given an intradermal injection of 25 μg of purified protein derivative of tuberculin. The following table gives the skin reactivity diameter of erythema and induration (in mm) for the two groups of rats.

Hypothesis Testing: Two-Sample Inference

Normal Populations with Equal Variances

Exercise

18% Protein Diet	*5% Protein Diet*
13.3	5.1
16.3	8.7
9.9	8.7
9.3	8.5
16.1	8.1
9.7	6.9
9.7	6.9
14.1	12.3

Assume that the two populations are normal with equal variances.

Hypothesis Testing: Two-Sample Inference
Normal Populations with Equal Variances

Exercise

(a) Test the hypothesis
$H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 > \mu_2$, *where* μ_1 *is the average level of tuberculin reactivity in the rats with a normal diet, and* μ_2 *is the average level of tuberculin reactivity in the malnourished rats. State your conclusion.*
(b) Construct a 95% confidence interval for $\mu_1 - \mu_2$. *Can we say that the skin reactivity diameter in the malnourished rats is at least 7mm smaller than in the control group?*

Hypothesis Testing: Two-Sample Inference
Normal Populations with Equal Variances

Solution

In what follows, Assume that the two populations are normal with equal variances.

(a) Here is the summary of the data:

size	*mean*	*variance*	*standard deviation*
$n_1 = 8$	$\bar{x}_1 = 12.3$	$s_1^2 = 8.994$	$s_1 = 2.99$
$n_2 = 8$	$\bar{x}_2 = 8.15$	$s_2^2 = 4.34$	$s_2 = 2.08$

Hypothesis Testing: Two-Sample Inference
Normal Populations with Equal Variances

Solution

The pooled variance and standard deviation are:

$$s_p^2 = \frac{7(8.994) + 7(4.34)}{14} = 6.667, \quad s_p = 2.582$$

The observed test statistic is:

$$t_0 = \frac{12.3 - 8.15}{2.582\sqrt{1/8 + 1/8}} = \frac{4.15}{(2.58)(0.5)} = 3.21$$

The p-value is $P(T > 3.21) < 0.005$ *since* $P(T > 2.997) = 0.005$, *where* T *has a* T_{14} *distribution.*

Hypothesis Testing: Two-Sample Inference

Normal Populations with Equal Variances

Solution

We reject H_0 and conclude that we have statistical evidence malnutrition diminishes tuberculin skin reactivity.

(c) We use $t = 2.145$. The interval is:

$$
\begin{aligned}
I_{\mu_1 - \mu_2} &= [4.15 \pm (2.145)(2.582)\sqrt{1/8 + 1/8}] \\
&= [4.15 - 2.77; 4.15 - 2.77] \\
&= [1.38; 6.92]
\end{aligned}
$$

We are 95% confident that the difference between $\mu_1 - \mu_2$ is between 1.38 and 6.92. We cannot say that the difference is larger than 7mm.

Hypothesis Testing: Two-Sample Inference
Normal Populations with Unequal Variances

Let $(X_1, X_2, \ldots, X_n)$ be a sample from $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
Let $(Y_1, Y_2, \ldots, Y_m)$ be a sample from $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.
Further, assume that $X \perp\!\!\!\perp Y$ and $\sigma_1^2 \neq \sigma_2^2$, both unknown.
Consider performing one of the following tests:

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2$$
$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 > \mu_2$$
$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 < \mu_2,$$

with a significance level α. Under H_0, the test statistic

$$T_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \approx T_\nu,$$

Hypothesis Testing: Two-Sample Inference
Normal Populations with Unequal Variances

where ν is called **Welch's number of degrees of freedom**:

$$\nu = \frac{(s_1^2/n + s_2^2/m)^2}{\frac{(s_1^2/n)^2}{n-1} + \frac{(s_2^2/m)^2}{m-1}}$$

should be round down to the nearest integer. Let t_0 denote the observed value of T_0.

- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, then reject H_0 if $|t_0| > t_{\nu;\alpha/2}$ or p-value$= 2P\left(T_0 > |t_0|\right) < \alpha$.
- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$, then reject H_0 if $t_0 > t_{\nu;\alpha}$ or p-value$= P\left(T_0 > t_0\right) < \alpha$.
- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$, then reject H_0 if $t_0 < t_{\nu;\alpha}$ or p-value$= P\left(T_0 < t_0\right) < \alpha$.

Exercise

There is a claim that the average blood volume in millimeters for males who are paraplegic and participate in vigorous physical activities is higher that that of males who are able-bodied and participate in normal physical activities. A sample of size $n_1 = 7$ *paraplegic men gives* $\bar{x}_1 = 1511.714$ *and* $s_1^2 = 49669.905$*. A sample of size* $n_2 = 10$ *able-bodied men gives* $\bar{x}_2 = 1118.400$ *and* $s_2^2 = 15297.600$*. Justify this claim using a* 95% *confidence interval for* $\mu_1 - \mu_2$*. Justify the claim using the method of hypothesis testing. Assume that the populations are normal with unequal variances.*

Solution

1) The number of degrees of freedom for T is

$$\nu = \frac{[49669.905/7 + 15297.600/10]^2}{[49669.905/7]^2/6 + [15297.600/10]^2/9} = 8.599$$

We round this number down *to* $\nu = 8$. *Since* $1 - \alpha = 0.95$, *then* $\alpha/2 = 0.025$. *Therefore,* $P(T \geq t_{0.025,8}) = 0.025$ *implies that* $t_{0.025,8} = 2.306$. *A* 95% *confidence interval is*

$$I_{\mu_1 - \mu_2} = \left[(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2;\nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

$$= [179.148;\ 607.480].$$

Solution (Cont'd)

Since this interval contains only positive values, we are 95% confident that μ_1 is larger than μ_2.

2) The statistical problem consists of testing the following:

$$H_0 : \mu_1 = \mu_2 \textit{ versus } H_1 : \mu_1 > \mu_2$$

The observed value of the test statistic, $T_0 \sim T_8$, is

$$t_0 = \frac{1511.714 - 1118.400}{\sqrt{49669.905/7 + 15297.600/10}} = 4.234$$

Since $4.234 = t_0 > t_{8;0.05} = 1.860$, then we reject H_0.

Hypothesis Testing: Two-Sample Inference

Comparing Two Proportions

- Let p_1 be the proportion of a population with a certain characteristic X and $(X_1, X_2, \ldots, X_n)$, $(n > 30)$, a large sample drawn from this population $(n > 30)$. Let $\hat{X}$ be the number of individuals with that characteristic in this sample and $\hat{P}_1 = \tilde{X}/n$, the corresponding sample proportion.
- Let p_2 be the proportion of a population with a certain characteristic Y and $(Y_1, Y_2, \ldots, Y_m)$, $(m > 30)$, a large sample drawn from this population. Let $\hat{Y}$ be the number of individuals with that characteristic in this sample and $\hat{P}_2 = \tilde{Y}/m$, the corresponding sample proportion.

Hypothesis Testing: Two-Sample Inference

Comparing Two Proportions

Consider performing one of the following tests:

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2$$
$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 > p_2$$
$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 < p_2,$$

with a significance level α. Under $H_0 : p_1 = p_2 = p$, the Central Limit Theorem yields that the test statistic

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n} + \frac{1}{m}\right)}} \approx \mathcal{N}(0, 1),$$

Hypothesis Testing: Two-Sample Inference

Comparing Two Proportions

where the **pooled sample proportion** estimate $\hat{p}$ is defined as follows:

$$\hat{p} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m} = \frac{\tilde{x} + \tilde{y}}{n+m}$$

and $\hat{q} = 1 - \hat{p}$. This approximation is decent when $n\hat{p}\hat{q} > 5$ and $m\hat{p}\hat{q} > 5$. Let z_0 denote the observed value of Z_0.

- If $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$, then reject H_0 if $|z_0| > z_{\alpha/2}$ or p-value$= 2P\left(Z_0 > |z_0|\right) < \alpha$.
- If $H_0 : p_1 = p_2$ versus $H_1 : p_1 > p_2$, then reject H_0 if $z_0 > z_\alpha$ or p-value$= P\left(Z_0 > z_0\right) < \alpha$.
- If $H_0 : p_1 = p_2$ versus $H_1 : p_1 < p_2$, then reject H_0 if $z_0 < -z_\alpha$ or p-value$= P\left(Z_0 < z_0\right) < \alpha$.

Exercise

A study looked at the effects of OC use on heart disease in women 40 to 44 years of age. The researchers found that among 100 current OC users at baseline, 27 women developed a myocardial infarction (MI) over a 3-year period, whereas among 100 non-OC users, 19 developed an MI over a 3-year period. Compare these two proportions from the point of view of hypothesis testing. Is there enough evidence that they differ from each other? (Use the significance level $\alpha = 0.05$*).*

Solution

Let p_1 denote the proportion of OC users that develop an MI and p_2, the proportion of non-OC users that develop an MI. The statistical problem consists of testing the following: $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$, with significance level $\alpha = 0.05$. It follows from assumptions that $\hat{p}_1 = 27/100 = 0.27$ and $\hat{p}_2 = 19/100 = 0.19$. Under $H_0 : p_1 = p_2 = p$, the Central Limit Theorem yields that the test statistic

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n} + \frac{1}{m}\right)}} \approx \mathcal{N}(0,1),$$

where the pooled sample proportion estimate $\hat{p}$ is calculated as follows:

Solution

$$\hat{p} = \frac{27+19}{100+100} = 0.23\ .$$

Notice that the aforementioned approximation is decent as $n\hat{p}\hat{q} = 17.71 > 5$ *and* $m\hat{p}\hat{q} = 15.39 > 5$. *The observed value of the test statistic* Z_0 *is*

$$z_0 = \frac{0.27 - 0.19}{\sqrt{(0.23)(0.77)\left(\frac{1}{100} + \frac{1}{100}\right)}} = 1.344.$$

Since p-value$= 2P(Z_0 > 1.344) = 2(0.09) = 0.182 > 0.05$, *then we fail to reject the null hypothesis* H_0.

Exercise

We want to compare the germination rate of a new strain of a plant against an old strain of the same plant. Below are the data.

	Germinated	Did Not Germinate	Total
Old Strain	125	15	140
New Stain	152	8	160

(a) Can we conclude that the germination rates differ? Use the level $\alpha = 0.05$.

(b) Construct a 98% confidence interval for the difference between the germination rates.

Solution

Let p_1 and p_2 be the germination rate for the old strain and the new strain, respectively. The sample germination rates are $\hat{p}_1 = 125/140 = 0.8929$ and $\hat{p}_2 = 152/160 = 0.95$. The pooled germination rate is
$\hat{p} = (125 + 152)/(140 + 160) = 0.9233.$
(a) We want to test $H_0 : p_1 = p_2$ against $H_1 : p_1 \neq p_2$. The observed value of the test statistic is

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{1/n_1 + 1/n_2}} = \frac{0.8929 - 0.95}{\sqrt{(0.9233)(1-0.9233)}\sqrt{1/140 + 1/160}} = -1.86.$$

Solution (Cont'd)

The p-value is $2\,P(Z > |-1.86|) = 2\,(1 - 0.9686) = 0.0628$. *Since the p-value is larger than the level of significance of* $\alpha = 5\%$, *then we should not reject* H_0. *There is not enough evidence to conclude that the germination rates are different.*

(b) A 98% confidence interval for $p_1 - p_2$ *is*

$$I_{p_1-p_2} = \left[\hat{p}_1 - \hat{p}_2 \pm 2.33\sqrt{\frac{\hat{p}_1\,(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2\,(1-\hat{p}_2)}{n_2}}\right]$$

$$= [-0.130, 0.016].$$

Hypothesis Testing: Two-Sample Inference
Paired Samples

Let (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) be the paired observations made on n individuals before and after a certain treatment (or when using two different treatments).
Let $(X_1, X_2, \ldots, X_n)$ be a sample from a population whose mean is μ_X and $(Y_1, Y_2, \ldots, Y_n)$ be a sample from a population whose mean is μ_Y. To compare μ_X and μ_Y,we consider performing one of the following hypothesis tests:

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2$$
$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 > \mu_2$$
$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 < \mu_2,$$

For this, define $D_i = X_i - Y_i, \quad i = \overline{1, n} \quad (n < 30)$ and assume that $D_i \sim \mathcal{N}(\mu_d, \sigma_d^2)$.

Hypothesis Testing: Two-Sample Inference

Paired Samples

with a significance level α. Under H_0, the test statistic

$$T_0 = \frac{\bar{D}}{S_d/\sqrt{n}} \sim T_{n-1},$$

where $\bar{D}$ and S_d^2 stand for the sample mean and variance of the random sample $D_i = X_i - Y_i$, $i = \overline{1, n}$; respectively. Let t_0 denote the observed value of T_0.

- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, then reject H_0 if $|t_0| > t_{n-1;\alpha/2}$ or p-value$= 2P(T_0 > |t_0|) < \alpha$.
- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$, then reject H_0 if $t_0 > t_{n-1;\alpha}$ or p-value$= P(T_0 > t_0) < \alpha$.
- If $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$, then reject H_0 if $t_0 < -t_{n-1;\alpha}$ or p-value$= P(T_0 < t_0) < \alpha$.

Exercise

The compound m-chloro-phenyl-piperazine (mCPP) is thought to affect appetite and food intake. Nine women took mCPP for two weeks (session 1), then took nothing for two weeks, and then took a placebo for two weeks (session 2). The weight loss (in kg) for each woman was recorded after each of the two sessions. Here is the table of the data set: (column 1-mCPP, column 2-placebo, column 3-difference)

Exercise (Cont'd)

Session 1 (x)	Session 2 (y)	Difference $d = x - y$
1.1	0.0	1.1
1.3	-0.3	1.6
1.0	0.6	0.4
1.7	0.3	1.4
1.4	-0.7	2.1
0.1	-0.2	0.3
0.5	0.6	-0.1
1.6	0.9	0.7
-0.5	0.2	-0.7

Is there enough evidence that the average weight loss for the first session larger than for the second session? Use the significance level $\alpha = 0.05$.

Solution

Let μ_1 denote the average loss for the first session and μ_2, for the second session. Let $\mu_d = \mu_1 - \mu_2$. This statistical problem consists of testing $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 > \mu_2$, with significance level $\alpha = 0.05$. From the data set, it follows that $\bar{d} = 0.76$ and $s_d = 0.88$. Under H_0,

$$T_0 = \frac{\bar{D}}{S_d/\sqrt{n}} \sim T_8.$$

The observed value of the test statistic T_0 is

$$t_0 = \frac{0.76}{0.88/\sqrt{9}} = 2.59.$$

Since $t_{8,0.05} = 1.860 < t_0 = 2.59$, then we do reject H_0.

Exercise

Initial Benzene Level (x_i)	Benzene Level after 3 Days (y_i)
28.4	27.4
27.3	26.3
25.5	25.6
29.4	24.5
30.2	28.7
31.3	29.6
28.6	27.5
28.4	28.4
26.5	23.2
27.3	24.3

Exercise

Is there any evidence that this species of indoor plants is effective in removing the benzene from the indoor air? Justify your answer using a 95% confidence interval and a test of hypothesis. Assume that the differences $D_i = X_i - Y_i, i = 1, \ldots, 10$ *satisfy the normality assumption.*

Solution

Assume that the differences $D_i = X_i - Y_i, i = 1, \ldots, 10$ satisfy the normality assumption.
Here is the summary of the data:

	Initial	*After 3 days*	*Difference*
Mean	$\bar{x} = 28.29$	$\bar{y} = 26.56$	$\bar{d} = 1.73$
Standard deviation	$s_x = 1.732$	$s_y = 2.102$	$s_d = 1.552$

Solution

The 95% confidence interval for

$$1.73 \pm 2.262\left(\frac{1.552}{\sqrt{10}}\right) = 1.73 \pm 1.11 = [0.62; 2.84]$$

where we used the fact that $P(T > 2.262) = 0.025$ *and* T *has a* $T(9)$ *distribution. Since the interval contains only positive values, we are 95% confident that* $\mu_X > \mu_Y$, *i.e. the average benzene level has been reduced.*

We test $H_0 : \mu_X = \mu_Y$ *against* $H_1 : \mu_X > \mu_Y$. *The observed value of the test statistic is:*

$$\frac{1.73}{1.552/\sqrt{10}} = 3.52$$

The p-value=$P(T > 3.52)$ *is between 0.001 and 0.005. Since the p-value is small, we reject* H_0, *and conclude that there is evidence that* $\mu_X > \mu_Y$.