Chapter 8
Hypothesis Testing
t Test

Outline

Introduction

The z-score is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

taken under the assumption that the population standard deviation is known.
If we are forced to replace the unknown σ^2 with its unbiased estimator s^2, then the statistic is known as t:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

We have previously noted that for independent normal random variables the distribution of the t statistic can be determined exactly and we used the t distribution to construct a confidence interval for the population mean μ.

We now turn to using the t-statistic as a test statistic for hypothesis tests of the population mean. As with several other procedures we have seen, the two-sided t test is a likelihood ratio test.

Guidelines for Using the t Procedures

- Except in the case of small samples, the assumption that the data are a simple random sample from the population of interest is more important that the population distribution is normal.
- For sample sizes less than 15, use t procedures if the data are close to normal.
- For sample sizes at least 15 use t procedures except in the presence of outliers or strong skewness.
- The t procedures can be used even for clearly skewed distributions when the sample size is large, typically over 40 observations.

These criteria are designed to ensure that $\bar{x}$ is a sample from a nearly normal distribution. When these guidelines fail to be satisfied, then we can turn to alternatives that are not based on the central limit theorem, but rather use the rankings of the data.

One-Sample Tests

The two-sided hypothesis

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0,$$

based on independent normal observations $X_1, \ldots, X_n$ with unknown mean μ and unknown variance σ^2 is a likelihood ratio test. The parameter space and null hypothesis space, are, respectively,

$$\Theta = \{(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma^2 > 0\} \quad \text{and} \quad \Theta_0 = \{(\mu, \sigma^2); \mu = \mu_0, \sigma^2 > 0\}.$$

The critical region is a level set for the t-statistic, $T(\mathbf{x})$ from the data $\mathbf{x}$.

$$C = \{|T(\mathbf{x})| > t_{n-1,\alpha/2}\}.$$

where $t_{n-1,\alpha/2}$ is the upper $\alpha/2$ tail probability of the t distribution with $n-1$ degrees of freedom.

The t-test

To show that the critical region is a likelihood ratio test, consider

$$\Lambda(x) = \frac{L(\mu_0, \hat{\sigma}_0^2|\mathbf{x})}{L(\hat{\mu}, \hat{\sigma}^2|\mathbf{x})},$$

we begin with independent normal observations $X_1, \ldots, X_n$ with unknown mean μ and *unknown* variance σ^2. The likelihood function

$$L(\mu, \sigma^2|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

We have seen that the maximum over the parameter set Θ occurs for

$$\hat{\mu} = \bar{x} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and over Θ_0 for

$$\hat{\mu}_0 = \mu \quad \text{and} \quad \hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

The t-test

Substituting back into the likelihood function,

$$L(\mu_0, \hat{\sigma}_0^2|\mathbf{x}) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} \exp -\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^{n}(x_i - \mu_0)^2 = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} \exp -\frac{2}{n},$$

$$L(\hat{\mu}, \hat{\sigma}^2|\mathbf{x})) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}}, \exp -\frac{2}{n},$$

and the likelihood ratio is

$$\Lambda(\mathbf{x}) = \frac{L(\mu_0, \hat{\sigma}_0^2|\mathbf{x})}{L(\hat{\mu}, \hat{\sigma}^2|\mathbf{x})} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{-n/2} = \left(\frac{\sum_{i=1}^{n}(x_i - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{-n/2}$$

The t-test

The critical region $\lambda(\mathbf{x}) \leq \lambda_0$ is equivalent to the fraction in parenthesis above being sufficiently large. In other words for some value c_α,

$$\begin{aligned} c_\alpha &\leq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n \left((x_i - \bar{x}) + (\bar{x} - \mu_0)\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + 2\sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \mu_0) + \sum_{i=1}^n (\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = 1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Continuing we find that

$$(c_\alpha - 1)(n-1) \leq \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2/(n-1)} = \frac{(\bar{x} - \mu_0)^2}{s^2/n}.$$

The t-test

$$(c_\alpha - 1)(n-1) \le T(\mathbf{x})^2$$

where

$$T(\mathbf{x}) = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

and s is the square root of the unbiased estimator of the variance.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Taking square roots, we have the critical region

$$C = \left\{\mathbf{x};\, \sqrt{(c_\alpha - 1)(n-1)} \le |T(\mathbf{x})|\right\}$$

Thus, we take $\sqrt{(c_\alpha - 1)(n-1)} = t_{n-1,\alpha/2}$.

One-Sample Tests

- Radon is formed as part of the normal radioactive decay chain of uranium.
- It is one of the densest substances that remains a gas under normal conditions.
- Radon gas from natural sources can accumulate in buildings, especially in confined areas such as attics, and basements.
- Epidemiological evidence shows a clear link between breathing high concentrations of radon and incidence of lung cancer.
    - According to the United States Environmental Protection Agency, radon is the second most frequent cause of lung cancer, after cigarette smoking, causing 21,000 lung cancer deaths per year in the United States.

One-Sample Tests

To check the reliability of radon detector, a university placed 12 detectors in a chamber having 105 picocuries of radon. (1 picocurie is 3.7×10^{-2} decays per second. This is roughly the activity of 1 picogram of radium 226.)

The two-sided hypothesis

$$H_0 : \mu = 105 \quad \text{versus} \quad H_1 : \mu \neq 105,$$

where μ is the mean value of the radon detectors. In other words, we are checking to see if the detector is biased either upward or downward.

The detector readings were:

91.9 97.8 111.4 122.3 105.4 95.0 103.8 99.6 96.6 119.3 104.8 101.7

One-Sample Tests

Using R, we find for an $\alpha = 0.05$ level significance test:

```
> mean(radon);sd(radon)
[1] 104.1333
[1] 9.397421
> qt(0.975,11)
[1] 2.200985
```

The t-statistic is

$$t = \frac{105 - 104.1333}{9.39742/\sqrt{12}} = -0.3195.$$

Thus, for a 5% significance test, $|t| < 2.200985$, the critical value and we *fail* to reject H_0.

One-Sample Tests

R handles this procedure easily.

```
> t.test(radon,alternative=c("two.sided"),mu=105)
One Sample t-test
data:  radon
t = -0.3195, df = 11, p-value = 0.7554
alternative hypothesis: true mean is not equal to 105
```

Power analyses are based on the noncentral t-distribution. Recall that for $Z_1, Z_2, \ldots Z_n$, independent standard normal random variables with standard deviation s_Z,

$$\tilde{T} = \frac{\sqrt{n}\bar{Z} + a}{s_Z}$$

has a t distribution with $n-1$ degrees of freedom and non-centrality parameter a.

Power Analysis

If we reproduce the calculation we made in determining power for a one-sample two-sided z-test, then, in the denominator, we add and subtract the mean μ to obtain

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{(\bar{X}-\mu)+(\mu-\mu_0)}{s/\sqrt{n}}$$

If the X_i have common standard deviation σ, we standardize the variables, writing

$$Z_i = \frac{X_i-\mu}{\sigma}.$$

Thus, $s_Z = s/\sigma$ is the standard deviation of the Z_i and upon dividing each term by σ,

$$\begin{aligned}\frac{(\bar{X}-\mu)+(\mu-\mu_0)}{s/\sqrt{n}} &= \frac{\sqrt{n}((\bar{X}-\mu)/\sigma+(\mu-\mu_0)/\sigma)}{s/\sigma} \\ &= \frac{\sqrt{n}\bar{Z}-\sqrt{n}(\mu_0-\mu)/\sigma}{s_Z}\end{aligned}$$

which has t-distribution with non-centrality parameter $a = \sqrt{n}(\mu-\mu_0)/\sigma$.

Power Analysis

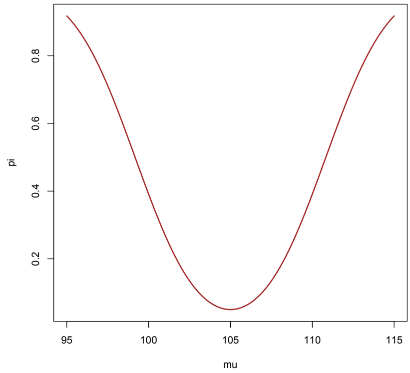
To plot the power function, π, we first enter the data.

```
> radon<-c(91.9,97.8,111.4,122.3,105.4,95.0,
  103.8,99.6,96.6,119.3,104.8,101.7)
> mu0<-105
> mu<-seq(95,115,length=101)
```

and estimate the standard deviation from the sample.

```
> a<-(mu0-mu)/(sd(radon)/sqrt(length(radon)))

> tstar<-qt(0.975,11)
> pi<-1-(pt(tstar,11,a)-pt(-tstar,11,a))
>plot(mu,pi,type="l",col="brown",lwd=2)
```



Power Analysis

The `power.t.test` command considers five issues

- the sample size `n`,
- the difference between the null and a fixed value of the alternative `delta`,
- the standard deviation s,
- the significance level α, and
- the power $1 - \beta$.

We can use `power.t.test` to drop out any one of these five and use the remaining four to determine the remaining value. For example, if we want to assure an 80% power against an alternative difference of 5 piconewtons,

```
> power.t.test(power=0.80,delta=5,sd=sd(radon),type=c("one.sample"))
```

The output shows that we need to make 30 measurements.

Power Analysis

Exercise. Fill in the tables based on the radon detector data set.

1. Consider a significance level $\alpha = 0.05$ and the standard deviation s from the data. Find the necessary number of observations.
2. Consider a significance level $\alpha = 0.05$ and a difference delta of 5 piconewtons. Find the power.

	power		
delta	0.7	0.8	0.9
5			
10			

	observations		
s	10	20	30
5.00			
9.40			

3. Which two values in each table are most extreme? Explain why.

Tests for Population Means

- Use the z-statistic when the standard deviations are known.
- Use the t-statistic when the standard deviations are computed from the data.

	null hypothesis	
t or z-procedure	one-sided	two-sided
single sample	$H_0 : \mu \leq \mu_0$ $H_0 : \mu \geq \mu_0$	$H_0 : \mu = \mu_0$
two samples	$H_0 : \mu_1 \leq \mu_2$ $H_0 : \mu_1 \geq \mu_2$	$H_0 : \mu_1 = \mu_2$

Tests for Population Means

The test statistic,

$$t = \frac{\text{estimate} - \text{parameter}}{\text{standard error}}.$$

t-procedure	parameter	estimate	standard error	degrees of freedom
one sample	μ	$\bar{x}$	$\frac{s}{\sqrt{n}}$	$n-1$
two sample	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	ν in W-S equation
pooled two sample	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$n_1 + n_2 - 2$

For one-sample and two-sample z procedures, replace the values s with σ and s_1 and s_2 with σ_1 and σ_2, respectively. Use the normal distribution for these tests.

Tests for Population Means

For a two-sample t-test, the test statistic does *not* have a t-distribution. The Behrens-Fisher problem is the problem of hypothesis testing concerning the difference between the means of two independent normally distributed populations when the variances of the two populations are *not* assumed to be equal.

One approach solves this by approximating a t distribution with the effective degrees of freedom ν calculated using the Welch-Satterthwaite equation:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

We have that

$$\min\{n_1, n_2\} - 1 \leq \nu \leq n_1 + n + 2 - 2.$$

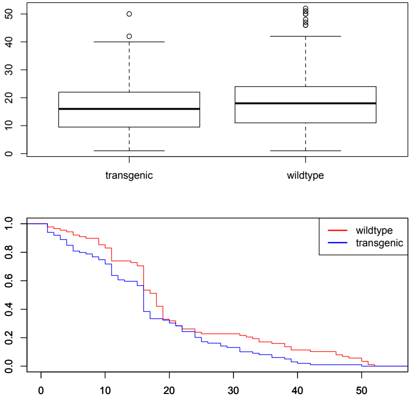
The value ν is determined by approximating a linear combination of squares of normal using a χ_ν^2 distribution.

Mosquito Life Span

Anopheles mosquitoes are the carrier of parasitic protozoans of the genus *Plasmodium*. The blood obtained from a bite from a female mosquito is used as a source of protein for the production of eggs. In this way infracted mosquitoes transmit malaria to humans.

We test to see if overstimulation of the insulin signaling cascade in the midgut of transgenic mosquitoes reduces the μ_t, the mean life span of these transgenic mosquitoes from that of the wildtype μ_{wt}.

$$H_0 : \mu_{wt} \leq \mu_t \quad \text{versus} \quad H_1 : \mu_{wt} > \mu_t.$$



Boxplot and survival function of lifespan in days for transgenic and wildtype mosquitoes.

Mosquito Life Span

R easily handles the analysis.

```
> t.test(transgenic,wildtype,alternative = c("less"))

Welch Two Sample t-test

data:  transgenic and wildtype
t = -2.4106, df = 169.665, p-value = 0.008497
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -1.330591
sample estimates:
mean of x mean of y
 16.54545  20.78409
```

NB. the number of degrees of freedom $\nu = 169.665$.