Topic 18 - Statistical Hypothesis Testing

Statistics for Managers

June 3, 1999

Statistical hypothesis testing is used to make statistical inferences about the unknown parameter(s) of a population based on a random sample.

Null and Alternative Hypothesis

The Null Hypothesis denotes a value or values for the unknown parameter; which if true, the investigator would take no action.

$$H_0: \mu = \mu_0$$

The Alternate Hypothesis denotes a value or values for the unknown parameter; which if true, the investigator would take action.

$$\text{Two Sided } \; H_a: \mu \neq \mu_0$$
$$\text{One Sided } \; H_a: \mu > \mu_0$$
$$\text{One Sided } \; H_a: \mu < \mu_0$$

Probability of Type I and Type II Errors

Since inferences are based on a sample of data, they are prone to error.

Future samples of data may lead to contradicting conclusions.

The Type I error occurs when a true null hypothesis is rejected given the data.

I.e., you act when you should not.

The Type II error occurs when a true alternate hypothesis is rejected given the data.

I.e., you do not act when you should.

Statistical methods are used to control the probability of error of both types of errors.

The <u>Power</u> of the statistical test is 1-β.

$$\alpha = \Pr\left\{ \text{True } H_0 \text{ Rejected} \middle| \text{DATA} \right\}$$

$$\beta = \Pr\left\{ \text{True } H_a \text{ Rejected} \middle| \text{DATA} \right\}$$

One-Sided Hypothesis Example:

In a manufacturing plant, plastic sheathing is specified to be at least two mils thick by one of the many quality measures. Set up the null and alternative hypothesis for a quality monitoring system that ensures the desired level of quality.

Answer:

The machine operator would act by adjusting the extruder rollers on the machine only if the plastic sheathing was too thin.

$$\text{Null } H_0 : \mu = \mu_0$$

$$\text{Alternative } H_a : \mu < \mu_0$$

$$\mu_0 = 2 \; mils$$

Two-Sided Hypothesis Example:

In nuclear power plant, the cold start procedure consists of bringing the reactor to 35% of power, and then to 65% of power, before full operation; a process that may take 12 hours. At each stage, engineers take measurements of several critical reactor attributes. For example, if binding energy[1] **for a given fuel rod does not have a mean rate of 11.5 MeV at 35% power, then the reactor could cascade into a critical configuration and leak radiation at subsequent power levels. Set up the hypothesis for a decision system at the 35% power level stage.**

Answer

The plant operators would not continue to power up the reactor if the binding energy did not meet specification. The action to be taken would be to shut down.

$$\text{Null } H_0 : \mu = \mu_0$$

$$\text{Alternative } H_a : \mu \neq \mu_0$$

$$\mu_0 = 11.5$$

[1]Excerpted from *Compton's Interactive Encyclopedia*.

Known SD: Z-Statistic

In statistical procedures, data is summarized into a much smaller set of values than the total sample size.

In sampling from a normal distribution with <u>known standard deviation</u>, the data are reduced to the sample mean and the sample standard deviation.

The <u>z-statistic</u> is the z-score of sample mean standardized by the location parameter and the known standard error of the mean.

The location parameter is taken from the null hypothesized value for the population mean.

In repeated samples from a normal distribution, the sample of values of the z-statistic follows the normal distribution.

In other words, if $n \geq 30$, by the Central Limit Theorem, the sample of values of the z-statistic follows the normal distribution.

$$z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$$

$$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}$$

Z-Statistic Decision Rules for Probability of Type I Error Equal to α.

One-Sided Testing

Reject H_0: $\mu = \mu_0$ in favor of H_a: $\mu > \mu_0$ if $z \geq z_{\propto}$.

Reject H_0: $\mu = \mu_0$ in favor of H_a: $\mu < \mu_0$ if $z \leq -z_{\propto}$.

Two-Sided Testing

Reject H_0: $\mu = \mu_0$ in favor of H_a: $\mu \neq \mu_0$ if $|z| \geq z_{\propto/2}$.

One-Sided Example

In a manufacturing plant, plastic sheathing is specified to be at least 2 mils thick by one of the many quality measures. The thickness is known to be N(μ,4). The quality control system has a probability of type I error equal to .05. In 25 samples on one day, the mean thickness was .43 mils. What should the machine operator do?

Recall the hypothesis:

$$\text{Null } H_0 : \mu = \mu_0$$
$$\text{Alternative } H_a : \mu < \mu_0$$
$$\mu_0 = 2 \ mils$$

Decision Rule:

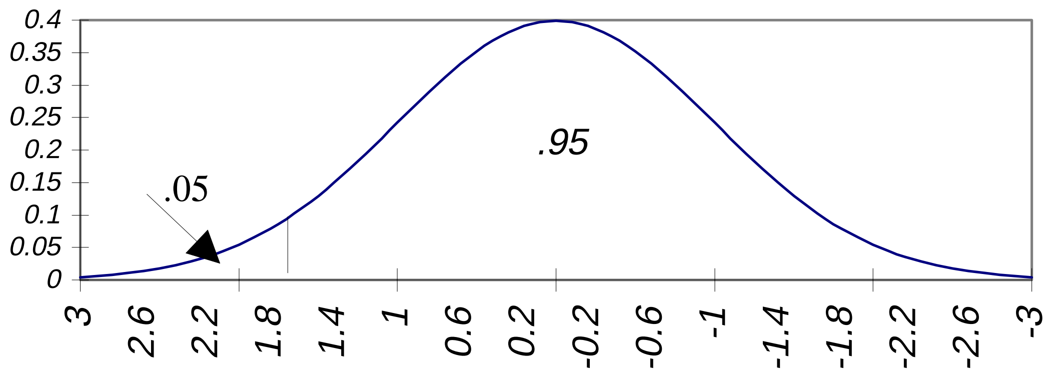
Reject H_0: $\mu = 2$ in favor of H_a: $\mu < 2$ if $z \leq -z_{.05} = -1.645$.

Calculation of Test Statistic:

$$z = \frac{.43 - 2}{.80} = -1.9625$$
$$\sigma_{\overline{X}} = \frac{4}{\sqrt{25}} = .80$$

Answer:

Reject the null hypothesis and shut down manufacturing to adjust machine.



Scale of *z*:

$-z_{.025} = -1.645$ 0

Scale of $\overline{X}$:

$2 - 1.645 \cdot 4/\sqrt{25}$

$= .6840$ $\mu_0 = 2$

Note:

The sample mean will fall in the interval (.6840, ∞) in 95% of the time during repeated sampling as long as the population mean is 2. However, it will fall below .6280, 5% percent of the time; resulting in needless shutting down of plastic sheathing manufacturing.

One-Sided Non-Normal Example:

In another manufacturing plant, set up a quality control system with α =0.01 using sample of size 9 for the null hypothesis with mean 2 and SD 4, and a one sided alternative with mean less than 2.

Answer:

We can't put together a system with any great precision using the z-test since the sample size is less than 30 and the population may not be normally distributed.

Two-Sided Example:

In nuclear power plant, the cold start procedure consists of bringing the reactor to 35% of power, and then to 65% of power, before full operation. At each stage, engineers make measurements of several critical reactor attributes. If the binding energy[2] does not have a mean rate of 11.5 MeV at 35% power, then the reactor could cascade into a critical configuration at subsequent power levels. Set up the hypothesis for a decision system at the 35% power level stage using a liberal α = .01. It is known that the population of measurement errors is normal with SD 1.5. On this day's power-up, the sample mean of nine observation is 10.1 MeV. What should the operators of the reactor do?

Recall the Hypothesis:

$$\text{Null } H_0 : \mu = \mu_0$$

$$\text{Alternative } H_a : \mu \neq \mu_0$$

$$\mu_0 = 11.5$$

Decision Rule:

Reject H_0: $\mu = 11.5$ in favor of H_a: $\mu \neq 11.5$ if $|z| \geq z_{.005} = 2.576$.

Calculation of Test Statistic:

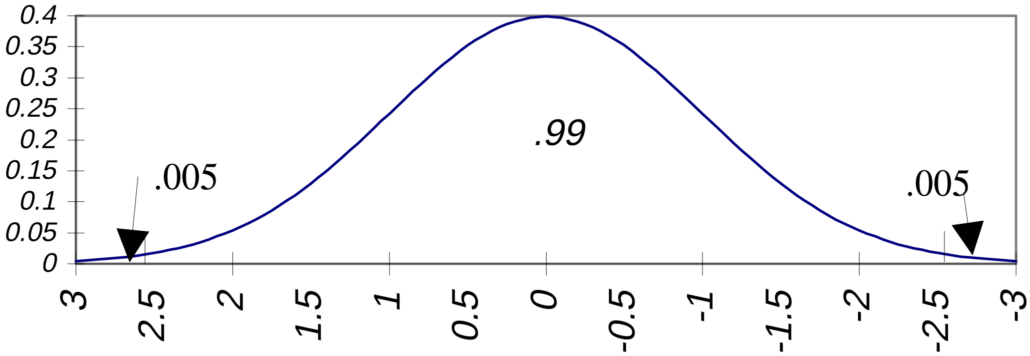
$$z = \frac{10.2 - 11.5}{.50} = -2.60$$

$$\sigma_{\overline{X}} = \frac{1.5}{\sqrt{9}} = .50$$

Answer:

Reject the null hypothesis and shut down the power-up.

2 Ibid.



Scale of *z*:

$-z_{.005} = -2.576$	0	$z_{.005} = 2.576$

Scale of $\overline{X}$:

$11.5 - 2.576 \cdot 1.5/\sqrt{9} = 10.2120$	**11.5**	$11.5 + 2.576 \cdot 1.5/\sqrt{9} = 12.7880$

Note:

The mean of a sample of size 9 will fall within (10.2120, 12.7880) in 99% of all random samples when sampling from N(11.5,1.5). However, it will fall outside this interval 1% of the time, resulting in needless shutting down of the reactor power-up. Better safe than sorry!

Unknown SD: Students t-Statistic

In statistical procedures, data is summarized into a much smaller set of values than the total sample size.

In sampling from a normal distribution without knowing the population standard deviation, the data are reduced to the sample mean and the sample standard deviation.

The Student's t-statistic is the sample mean standardized by the location parameter and the sample standard deviation.

The t-statistics has the Student's t-distribution with n-1 degrees of freedom.

The location parameter is taken from the null hypothesized value for the population mean.

$$t = \frac{\overline{X} - \mu_0}{s_{\overline{X}}}$$

$$s_{\overline{X}}^2 = \frac{s_X^2}{n}$$

t-Statistic Decision Rules for Probability of Type I Error Equal to α.

One-Sided Testing

Reject H_0: $\mu = \mu_0$ in favor of H_a: $\mu > \mu_0$ if $t \geq t_\alpha(n-1)$.

Reject H_0: $\mu = \mu_0$ in favor of H_a: $\mu < \mu_0$ if $t \leq -t_\alpha(n-1)$.

Two-Sided Testing

Reject H_0: $\mu = \mu_0$ in favor of H_a: $\mu \neq \mu_0$ if $|t| \geq t_\alpha(n-1)$.

One-Sided t Example

In an energy study, a utility company ran a campaign in a certain homogeneous community to decrease the amount of electricity used. Before the campaign, the average was 14,400-kilowatt hours per year. A sample of 25 households resulted in a mean of 14,052 with a sample standard deviation of 355-kilowatt hours. Was the advertising campaign successful?

The hypothesis:

$$\text{Null } \; H_0 : \mu = \mu_0$$

$$\text{Alternative } \; H_a : \mu < \mu_0$$

$$\mu_0 = 14{,}400 \; kwh$$

Decision Rule Using $\alpha = .05$

Reject H_0: $\mu = 14{,}400$ in favor of H_a: $\mu < 14{,}400$ if $t \le -t_{.05}(24) = -1.711$.

Calculation of Test Statistic:

$$t = \frac{14052 - 14400}{71} = -4.9014$$

$$s_{\overline{X}} = \frac{355}{\sqrt{25}} = 71$$

Answer:

The campaign was successful and we recommend that the electric utility extend it to other communities.

P-Values

The p-value denotes the probability of observing a larger value of the test statistic in future samples when the null hypothesis is true.

$$p = \Pr\left\{ t > t_{obs} \mid H_0, n-1 df \right\}$$

Equivalently, tests of hypotheses be based on the p-value than the actual test statistic.

$$\text{Reject } H_0 \text{ if}$$
$$p \le \alpha$$

The <u>two-tailed significance</u> is a p-value based on the absolute value of the t-statistic for testing the two-sided alternative hypotheses.

$$p = \Pr\left\{ |t| > t_{obs} \mid H_0, n-1 df \right\}$$

The test of two-sided hypotheses is carried out in the same way as the one-sided hypotheses.

$$\text{Reject } H_0 \text{ if}$$
$$p \le \alpha$$

Example Dataset: Westwood Company

Westwood manufactures standard wall clocks for non-industrial consumption. Wholesalers order the clocks in Lot Sizes. Westwood is repeating a study about how many Man-Hours it takes to manufacture recent Lot Sizes for costing purposes and for comparing the efficiency of its manufacturing process to historical records. Several years ago, when Westwood originally did this study, it was determined that it should take on average 2 man-hours to produce one clock. Management is concerned the production hours are now longer since new workers do not have good job skills.

Westwood Data

		Production Run i	*Lot Size X(i)*	*Man-Hours Y(i)*	*Rate Hours per Lot Size m(i)*
1		*1*	*30*	*73*	*2.43333*
2		*2*	*20*	*50*	*2.50000*
3		*3*	*60*	*128*	*2.13333*
4		*4*	*80*	*170*	*2.12500*
5		*5*	*40*	*87*	*2.17500*
6		*6*	*50*	*108*	*2.16000*
7		*7*	*60*	*135*	*2.25000*
8		*8*	*30*	*69*	*2.30000*
9		*9*	*70*	*148*	*2.11429*
10		*10*	*60*	*132*	*2.20000*
Total	*N*	*10*	*10*	*10*	*10*
	Mean	*5.50*	*50.00*	*110.00*	*2.23910*
	Variance	*9.167*	*377.778*	*1517.778*	*1.8E-02*
	Std. Deviation	*3.03*	*19.44*	*38.96*	*.13387*

Computer Results

One-Sample Test

	Test Value = 2					
					95% Confidence Interval of the Difference	
	t	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Lower*	*Upper*
Rate Hours per Lot Size m(i)	*5.648*	*9*	*.000*	*.2391*	*.1433*	*.3349*

One-Sample Test

	Test Value = 2.1					
					95% Confidence Interval of the Difference	
	t	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Lower*	*Upper*
Rate Hours per Lot Size m(i)	*3.286*	*9*	*.009*	*.1391*	*4.3E-02*	*.2349*

One-Sample Test

	Test Value = 2.2					
					95% Confidence Interval of the Difference	
	t	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Lower*	*Upper*
Rate Hours per Lot Size m(i)	*.923*	*9*	*.380*	*3.910E-02*	*-6.E-02*	*.1349*