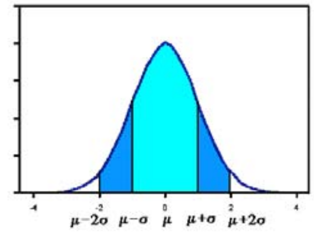
Sta108 Fall 2007 Utts



Normal Probability Distribution

Random variable X has a Normal distribution with mean μ and variance σ^2, or std.dev. σ.
This distribution curve is symmetric, center is the mean μ and spread is the std.dev. σ.
<u>Standard Normal Random Variable</u>:
If X = normal r.v. mean μ and std.dev. σ, then a transformation: $Z = \dfrac{X - \mu}{\sigma}$ (z-score) is standard normal r.v. with mean $\mu = 0$, std.dev. $\sigma = 1$, that is: $X \sim N(\mu, \sigma^2) \Rightarrow Z \sim N(0,1)$

Use normal tables to find probabilities for z (areas under curve). Normal tables give: $P(z < a)$, see shade in Tables picture.
Total Area under normal curve is = 1.
$P[x$ falls into interval from a to $b] = P(a < x < b)$ = area under the curve between a and b.
$P(a < x < b) = P(x < b) - P(x < a)$;
$P(x > a) = 1 - P(x \geq a) = 1 - P(x < a)$.

T-distribution

Similar to standard normal distribution, symmetric at 0, but has heavier tails.
A new parameter: degrees of freedom (*df*), determines the shape (how flat) of the distribution.
t-score will be different with every degree of freedom.
When df = 1 => t-distr. is flat. When df $\geq 30 \approx \infty$ => t-distr. is same as standard normal distr.

<u>Sampling distribution</u> of a statistic/estimator – the probability distribution for the possible values of the statistic that results when random samples of size n are repeatedly drawn from the population.

<u>*Central Limit Theorem*</u>: If random samples of n observations are drawn from a non-normal population with finite mean μ and standard deviation σ, then, when n is *large*, the sampling distribution of the sample mean $\bar{x}$ is approximately normally distributed, with mean μ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$. The approximation becomes more accurate as n becomes large.

Sampling Distribution of the <u>Sample Mean $\bar{x}$</u>

When a random sample of n measurements is selected from a population with mean μ and standard deviation σ =>
Sampling Distribution of the Sample Mean $\bar{x}$ has mean μ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$.

If $n \geq 30$, then $\bar{x}$ has approximately-Normal distribution with mean μ and standard error $SE(\bar{x}) = \dfrac{\sigma}{\sqrt{n}}$.

Standardized z-score: $z = \dfrac{\bar{x} - \mu}{SE(\bar{x})} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

(1-α)% Confidence Interval: (general formula)

[Point estimator] ± [Multiplier] × [Standard error of the estimator]

<u>(1-α)% CI for Population Mean μ</u>:

$\bar{x} \pm t_{\alpha/2} \dfrac{s}{\sqrt{n}}$, assumption: the sample is randomly selected from a normally distributed population.

Use t-distr with d.f. = n – 1. Why?

Because 1 degree of freedom is lost for estimating $\bar{x}$ in the formula of standard deviation: $s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n - 1}}$

Hypothesis Tests

1. Null and Alternative Hypotheses (Ho and Ha)
2. Test Statistic
3. Critical value / rejection region approach
4. Decision Rule (based on p-value), state: reject / do not reject Ho
5. Conclusion (Interpretation in context of the original problem)

Tests for Population Mean μ:

Type of test	Null and Alternatives	Test Statistic	p-value (use t-distr. table)
	Ho: $\mu = \mu_0$	$t^* = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	Rule: Reject Ho if p-value < α
Right-tailed test:	Ha: $\mu > \mu_0$		P(t > t*) d.f = n – 1
Left-tailed test:	Ha: $\mu < \mu_0$		P(t < t*)
Two-tailed test:	Ha: $\mu \neq \mu_0$		2 P(t > |t*|)

Check assumptions that the sample is random and sample comes from normal population.