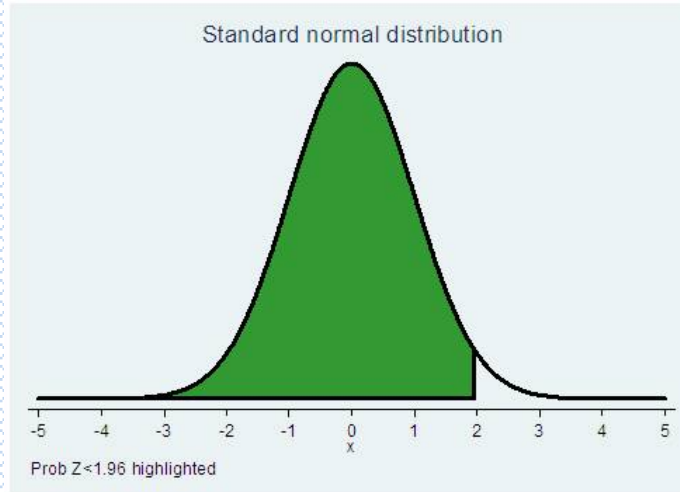
Confidence Intervals & Hypothesis Testing

- Briefly: Where are we & where are we going
- CLT and 95% confidence intervals
- Hypothesis testing in general
- Hypothesis testing of one mean

Review of the Normal Distribution



- For normal distribution
  - Stata calculates P(Z<z)
  - If you have a z-value and you want p [P(Z<z)], use

```
display normal(z)
di normal(1.96)
.9750021
```

  - If you have a z-value and you want p [P(Z>z)], use

```
display 1-normal(z)
di 1-normal(1.96)
.0249979
```

  - If you have p [P(Z<z)] and you want z-value, use

```
display invnormal(p)
. di invnormal(.975)
1.959964
```

  - If you have p [P(Z>z)] and you want z-value, use

```
display invnormal(1-p)
di invnormal(.025)
-1.959964
```

Or use what you know about the symmetry of the distribution

Review of Student's t Distribution

- For t distribution
  - Stata calculates P(T>t) for each d.f.
  - If you have a t value and you want p [P(T>t)], use

```
display ttail(df,t)
di ttail(99,1.96)
.02640356
```

  - If you have a t value and you want p [P(T<t)], use

```
display 1-ttail(df,t)
di 1-ttail(99,1.96)
.97359644
```

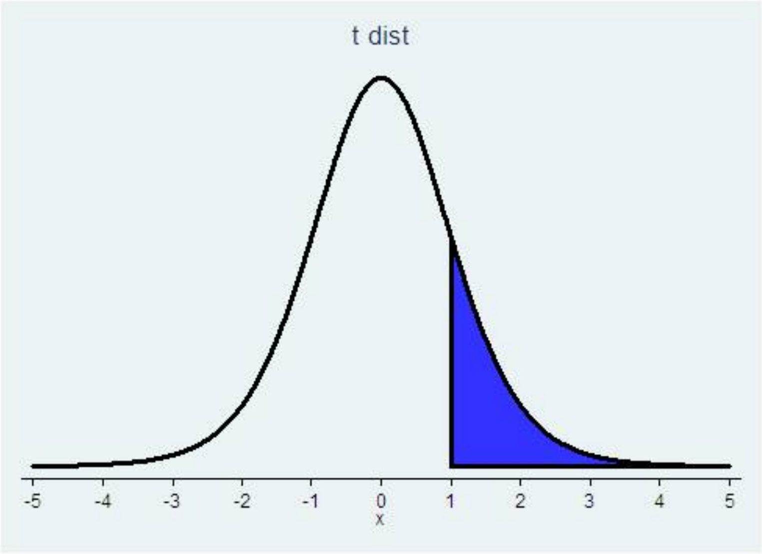
  - If you have p [P(T>t)] and you want t-value, use

```
display invttail(df,p)
di invttail(99,.025)
1.984217
```

  - If you have p [P(T<t)] and you want t value, use

```
display invttail(df,1-p)
di invttail(99,.975)
-1.984217
```

Or use what you know about the symmetry of the distribution



Confidence intervals

- Knowing the distribution of the sample means we write down

$$P(-1.96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96) = 0.95$$

and do the math to rearrange

$$P(\bar{X} - 1.96\sigma/\sqrt{n} \leq \mu \leq \bar{X} + 1.96\sigma/\sqrt{n}) = 0.95$$

- If you had drawn your sample 100 times, approximately 95 of the confidence intervals would include the true population mean μ.
  - *The interval is random.*
- These intervals can be derived for any point estimate, e.g. risk ratio, rate, etc. with known distribution.
- We use either z or t depending on whether the variance is known

Confidence intervals for <u>proportions</u>

- As n increases, the binomial distribution approaches the normal distribution
- Also, as n increases, the distribution of an estimated <u>proportion</u> approaches the normal distribution
  - $x/n = \hat{p} \sim N(\, p, \sqrt{p(1-p)/n}\,)$

    x/n is the proportion of successes

    p is the population proportion

    $\sqrt{p(1-p)/n}$ is the standard deviation of x/n if X is binomial

- So we can use the normal distribution to calculate confidence intervals for p

$$\hat{p} - 1.96 * \sqrt{\hat{p}(1-\hat{p})/n}\ ,\ \hat{p} + 1.96 * \sqrt{\hat{p}(1-\hat{p})/n}$$

Confidence intervals for <u>proportions</u>

- But if n is small or p is small or both then the normal approximation is not good and the intervals using the normal approximation are not good
- We need to use the binomial distribution
- Confidence intervals using the binomial distribution are called exact confidence intervals
- Exact confidence intervals are not symmetric around $\hat{p}$ (because the binomial distribution is not symmetric for small n or small p)

Hypothesis testing

- Confidence intervals tell us about the uncertainty in a point estimate of the population mean or proportion
- Hypothesis testing allows us to draw a conclusion about a population parameter that we have estimated in a sample
- Remember, we are taking samples from the population and we never know the truth.

some journals now ONLY want confidence intervals

Statistical hypothesis tests: cheat sheet

Data and comparison type	**Alternative hypotheses**	**Parametric test Stata command**
Numerical; One mean	H_a: $\mu \neq \mu_a$ (two-sided) H_a: $\mu > \mu_a$ or $\mu < \mu_a$ (one-sided)	Z or t-test • ttest var1==hypoth val.*
Numerical; Two means, <u>paired data</u>	H_a: $\mu_1 \neq \mu_2$ (two-sided) H_a: $\mu_1 > \mu_2$ or $\mu < \mu_a$ (one-sided)	Paired t-test • ttest var1==var2*
Numerical; Two means, independent data	H_a: $\mu_1 \neq \mu_2$ (two-sided) H_a: $\mu_1 > \mu_2$ or $\mu < \mu_a$ (one-sided)	T-test (equal or unequal variance) • ttest var1, by(byvar) • ttest var1, by(byvar) unequal
Numerical, Two or more means, independent data		
Dichotomous; One proportion	H_a: $p \neq p_a$ (two-sided) H_a: $p > p_a$ or $p < p_a$ (one-sided)	Proportion test • prtest var1=hypoth value* • bitest var1=hypoth value
Dichotomous; two proportions	H_a: $p_1 \neq p_2$ (two-sided) H_a: $p_1 > p_2$ (one-sided)	Proportion test (z-test) • prtest var1, by(byvar)

Hypothesis testing for a mean

- To conduct a hypothesis test about the mean of a population, we hypothesize a value for it, and call that value μ_0.
  - We write this $H_0 : \mu = \mu_0$
  - We call this the null hypothesis
- We set an alternative hypothesis for all other values of μ
  - We write this: $H_A : \mu \neq \mu_0$
  - We call this the alternative hypothesis
- We *set* the probability of incorrectly rejecting the null that we are willing to live with in advance, *before collecting the data and doing the analysis*
- That probability is called the *significance level*
- The significance level is denoted by α or alpha, and is frequently set to 0.05 or 5% error

Hypothesis testing

- In hypothesis testing, a true null hypothesis might be mistakenly rejected (if the sample you drew was very different from the null just by chance)
- This mistake is called a Type I error
  - The probability of a Type I error is chosen in advance
  - This probability is called the significance level, α

The probability of obtaining a mean at or more extreme than the observed mean, *given that the null hypothesis is true,* is the p-value of the test.

Basic steps in Hypothesis testing

- Before the test, set the significance level, α
  - This is the probability of rejecting the null hypothesis when it is true P(Type I error) that you can live with

- When you run the test, you get a probability of observing a sample mean as extreme as you did, given that the null hypothesis is true
  - This probability is the p-value

- If the p-value is less than α, (e.g. α=0.05 and p=0.013), then the test is called statistically significant and the null hypothesis is rejected

Hypothesis Testing

1. State the hypotheses

- H_0 (the null hypothesis) and H_A (the alternative hypothesis) must be mutually exclusive (no overlap) and include all the possibilities
- The alternative is the complement of the null

2. Determine the compatibility with the null hypothesis

- We determine if there is enough evidence to reject the null hypothesis (i.e. calculate the test statistic and find the p-value)
- If the null hypothesis is true, what is the probability of obtaining the sample data as extreme or more extreme?
    - Calculate a test statistic
    - Find the probability of the test statistic if the null is true
  - *This probability is called the p-value*

Hypothesis Testing

3. Reject or fail to reject

- If the p-value (the probability of obtaining the sample data and corresponding test statistic if the null hypothesis is true) is sufficiently small (we often use 5%) then we reject the null and say the test was statistically significant.
- This 5% is α, the significance level

- *<u>If we fail to reject the null hypothesis, it does not mean that we accept it</u>.*

Hypothesis Testing of one mean

- We want to test whether a mean is equal to an hypothesized value
- If we believe there might be deviations in either direction, we use a two-sided test
- Two sided test:
  - Null hypothesis: H_0: $\mu=\mu_0$
  - Alternative hypothesis: H_A: $\mu\neq\mu_0$
- If we only care about values above or below a certain value, we use a one-sided test
- One sided test:
  - Null hypothesis: H_0: $\mu\geq\mu_0$
  - Alternative hypothesis: H_A: $\mu<\mu_0$

or

  - Null hypothesis: H_0: $\mu\leq\mu_0$
  - Alternative hypothesis: H_A: $\mu>\mu_0$

Hypothesis Testing of one mean

- The distribution of the sample mean $\bar{x}$ if n is large enough is normal by the CLT. So you can calculate the z statistic:

$$z_{stat} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

If σ is not known, calculate

$$t_{stat} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Hypothesis Testing of one mean: one-sided

Non-pneumatic anti-shock garment for the treatment of obstetric hemorrhage in Nigeria

- Mean initial blood loss 1413.1 ml; SD=491.3; n=83
- Our question: Are these women sampled from a population that is hemorrhaging (>750 ml blood loss)?

Data adapted from Miller, S et al. Int J Gynecol Obstet (2009)

Hypothesis Testing of one mean: one-sided

- The null hypothesis is they are not hemorrhaging

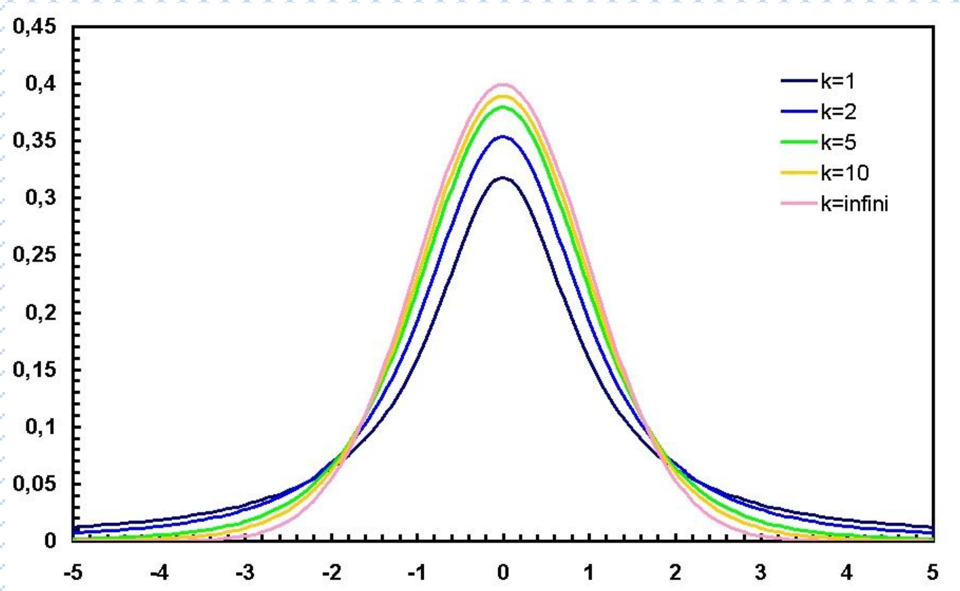
  H_0: μ≤750 H_A: μ>750 α=0.05

- T-statistic:
$$t_{stat} = \frac{\overline{X} - \mu_0}{s / \sqrt{n}} = \frac{1413.1 - 750}{491.3 / \sqrt{83}} = 12.3$$

- p-value: P(t>12.3) with 82 d.f.

Data adapted from Miller, S et al. Int J Gynecol Obstet (2009)

Hypothesis Testing of one mean



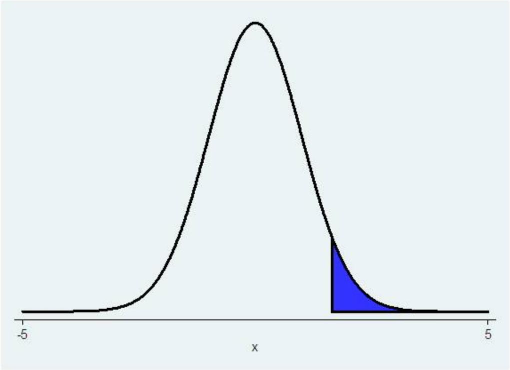
- 12.3 is off the graph

- So the probability of observing a sample mean of 1413 with n=83 and SD=431 if the true mean is <=750 is very very low

Hypothesis Testing of one mean: one-sided

P(T>test stat)

```
. display ttail(82, 12.3)
1.308e-20
```



The p-value is <0.001, which is a lot less than 0.05, therefore we reject the null

Hypothesis Testing of one mean: one-sided

Another way: Stata *calculate* code for ttests:

ttesti samplesize samplemean samplesd hypothesizedmean

ttesti 83 1413.1 491.3 750

```
One-sample t test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      83      1413.1    53.92718       491.3    1305.822    1520.378
------------------------------------------------------------------------------
    mean = mean(x)                                                t =  12.2962
    Ho: mean = 750                                   degrees of freedom =       82

   Ha: mean < 750              Ha: mean != 750              Ha: mean > 750
   Pr(T < t) = 1.0000      Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Stata gives you 3 alternative hypotheses

→ We reject the null hypothesis