

Biost 517

Applied Biostatistics I

.....
 Scott S. Emerson, M.D., Ph.D.
 Professor of Biostatistics
 University of Washington

Lecture 10: Inference About Means: One Sample

November 5, 2007

1

© 2002, 2003, 2005 Scott S. Emerson, M.D., Ph.D.

Lecture Outline

-
- Point Estimates
 - Confidence Intervals
 - Hypothesis Tests

2

Inference for Means

-
- Most common parameter used as a basis for statistical inference is the mean
 - Tends to reflect a wide variety of differences between distributions
 - E.g., extremely sensitive to changes in the tail of distributions
 - Statistical theory allow us to know the sampling distribution, and thus allows us to do inference

3

Point Estimate

-
- Most often estimate population mean with sample mean
 - Always unbiased estimate for the true mean
 - Tends to true mean across replicate experiments
 - Always consistent estimate for the true mean
 - Tends to true mean as sample size increases
 - Often minimum variability
 - Especially when the distribution of measurements is approximately normal

4

Approximate Sampling Distn

- Sample means are asymptotically normally distributed

Data (X_1, X_2, \dots, X_n) are

independent,

identically distributed,

with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$

For large n : $\bar{X} \sim N\left(\text{mean } \mu, \text{var } \frac{\sigma^2}{n}\right)$

5

What Should We Expect?

- If true mean is μ , true SD is σ , and sample size is n : sample mean should be

between $\mu - 1.645 \frac{\sigma}{\sqrt{n}}$ and $\mu + 1.645 \frac{\sigma}{\sqrt{n}}$ with probability 90%

$\mu - 1.96 \frac{\sigma}{\sqrt{n}}$ and $\mu + 1.96 \frac{\sigma}{\sqrt{n}}$ 95%

$\mu - 2.576 \frac{\sigma}{\sqrt{n}}$ and $\mu + 2.576 \frac{\sigma}{\sqrt{n}}$ 99%

$\mu - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ and $\mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ 100(1- α)%

6

Confidence Intervals

- Frequentist confidence interval estimates
 - “For what values of the population parameter are these data fairly typical?”
 - How should we define “typical”?
 - Not in the upper extreme of the sampling distn?
 - Not in the lower extreme of the sampling distn?
 - In neither of the “tails” of the sampling distn?
 - How should we define extreme?
 - 5%, 2.5%, 1%?

7

Lower Confidence Bound

- The true mean that is so low that we would not expect so high a sample mean
 - “not have expected” = probability is less than α

If $\mu = \mu_L$, with probability 100(1- α)% we expect

$$\bar{X} \leq \mu_L + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

So 100(1- α)% lower confidence bound is

$$\mu_L = \bar{X} - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

8

Upper Confidence Bound

- The true mean that is so high that we would not expect so low a sample mean
 - “not have expected” = probability is less than α

If $\mu = \mu_U$, with probability $100(1 - \alpha)\%$ we expect

$$\bar{X} \geq \mu_U - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

So $100(1 - \alpha)\%$ upper confidence bound is

$$\mu_U = \bar{X} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

9

100(1- α)% Confidence Interval

- Set of all true means that reasonably result in the observed sample mean
 - “reasonably” = central $100(1 - \alpha)\%$ of sampling distn

$100(1 - \alpha)\%$ confidence interval is (μ_L, μ_U)

$$\mu_L = \bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\mu_U = \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

10

Small Sample Adjustment

- Almost always have to estimate the population standard deviation
 - With continuous data, use the sample standard deviation s
- Use critical values for the t distribution
 - Exactly correct if the data were normal
 - Generally behaves well in other settings

11

100(1- α)% Confidence Interval

- Set of all true means that reasonably result in the observed sample mean
 - “reasonably” = central $100(1 - \alpha)\%$ of sampling distn

$100(1 - \alpha)\%$ confidence interval is (μ_L, μ_U)

$$\mu_L = \bar{X} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$$

$$\mu_U = \bar{X} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$$

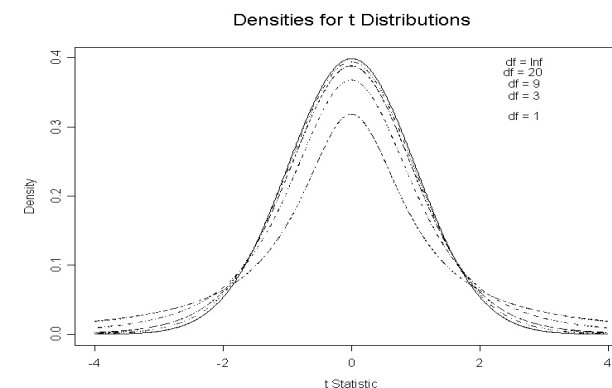
12

t Distribution

- Properties of the t distribution
 - “Degrees of freedom” related to sample size
 - Usually the sample size minus the number of parameters used to estimate the mean
 - Symmetric about 0
 - Heavier tails with decreasing degrees of freedom
 - 1 degree of freedom is Cauchy (no mean)
 - Infinite degrees of freedom is standard normal

13

t Distribution Densities



14

t Distribution Quantiles

- Selected upper quantiles of the t distribution: $t_{df, 1-\alpha}$

df	.01	.025	.05
1	31.821	12.706	6.314
3	4.541	3.182	2.353
9	2.821	2.262	1.833
20	2.528	2.086	1.725
50	2.403	2.009	1.676
Inf	2.326	1.960	1.645

15

Stata: CI for Population Mean

- `"ci var, level(#)"`
 - # is an integer between 10 and 99
 - a default can be set by "set level #"
- `"cii #n #mn #sd, level(#)"`
 - "immediate" CI by supplying n, sample mean, and sample standard deviation
- Commands for t test also give 95% CI
 - `"ttest var"`
 - `"ttesti #n #mn #sd #val"`

16

Example: FEV in Smokers

```
. bysort smoker: ci fev
```

```
-> smoker = 0
```

Var	Obs	Mean	Std. Err.	[95% Conf. Interval]
fev	589	2.57	0.04	2.50 2.63

```
-> smoker = 1
```

Var	Obs	Mean	Std. Err.	[95% Conf. Interval]
fev	65	3.28	0.09	3.09 3.46

17

Example: FEV in Smokers

- Best estimate of mean FEV
 - Nonsmokers: 2.57 l / sec
 - Smokers: 3.28 l / sec
- Interval estimate for mean FEV:
 - Nonsmokers: 95% confident that the true mean is between 2.50 l / sec and 2.63 l / sec
 - Smokers: 95% confident that the true mean is between 3.09 l / sec and 3.46 l / sec

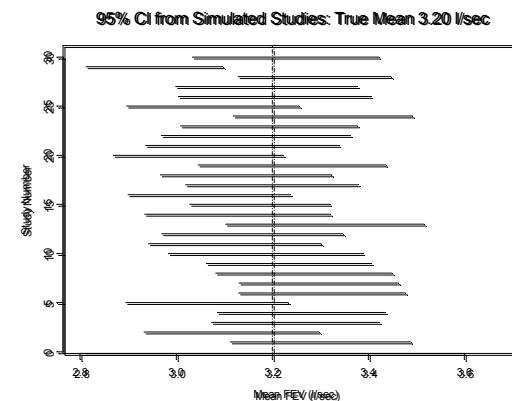
18

Interpretation of CI: Boring

- 95% CI for Smokers: 3.09 to 3.46 l / sec
- CORRECT, BUT BORING:
 - Of all CI computed in this manner, 95% of them will “cover” the true mean
 - (This says nothing about the numbers 3.09 and 3.46, because in repeated experiments, we would get different CI)

19

“Forest Plot” of Replicated CI



20

Interpretation of CI: Wrong

- 95% CI for Smokers: 3.09 to 3.46 l / sec
- **WRONG:**
 - Probability is 95% that the true mean is between 3.09 and 3.46 l / sec
 - (This is a Bayesian statement that would have to be based on a prior distribution for the mean FEV)

21

Interpretation of CI: Best

- 95% CI for Smokers: 3.09 to 3.46 l / sec
- **BEST:**
 - The sample mean we observed (3.28 l / sec) is typical of what we expect if the true mean were between 3.09 l / sec and 3.46 l / sec
 - (Frequentist confidence intervals are statements about the probability of observing data under hypothesized true values of the parameter)

22

Comparison of CI

- 95% CI for Nonsmokers: 2.59 to 2.63 l/sec
- 95% CI for Smokers: 3.09 to 3.46 l/sec
 - The two CI do not overlap
 - There is no true value of the mean that would typically lead to both of these observations
 - This finding ensures a “statistically significant” difference between the true means
 - BUT: Statistical significance can occur with overlapping confidence intervals

23

Hypothesis Tests of Means

- One sample t test: To test hypotheses that the true mean is in some particular range
 - Null Hypothesis
 - Usually status quo
 - We will tend to believe this unless we “prove” otherwise
 - What we hope to disprove
 - Alternative Hypothesis
 - What we hope is true

24

One- vs Two-sided tests

- One sided test of greater alternative
 - Null $H_0: \mu \leq \mu_0$ vs Alt : $\mu \geq \mu_1 > \mu_0$
- One sided test of lesser alternative
 - Null $H_0: \mu \geq \mu_0$ vs Alt : $\mu \leq \mu_1 < \mu_0$
- Two sided test
 - Null $H_0: \mu = \mu_0$ vs Alt : $\mu \neq \mu_0$
 - Lower Alt: $H_{1-}: \mu < - \mu_1$
 - Null $H_0: \mu = \mu_0$
 - Upper Alt: $H_{1+}: \mu > \mu_1$

25

Choosing One- vs Two-sided

- What will change your behavior?
 - E.g., New treatment vs Placebo
 - One-sided test
 - Only adopt new treatment if better
 - E.g., Existing treatment vs Placebo
 - Two-sided test
 - Push new treatment if better
 - Neutral if about equal
 - Warn against new treatment if worse

26

Computing P values

- Frequentist P value
 - Probability of observing as (or more) extreme results

Suppose we observe $\bar{X} = \bar{x}$

- Lower one - sided P value $\Pr(\bar{X} \leq \bar{x} \mid \mu = \mu_0)$
- Upper one - sided P value $\Pr(\bar{X} \geq \bar{x} \mid \mu = \mu_0)$
- Two - sided P value $\Pr(|\bar{X} - \mu| \geq |\bar{x} - \mu_0| \mid \mu = \mu_0)$

27

Computing P values using t

- Using t statistic
 - Standardized statistic $T = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \sim t_{n-1}$

Lower one - sided P value $\Pr\left(t_{n-1} \leq \frac{\bar{x} - \mu_0}{s / \sqrt{n}}\right)$

Upper one - sided P value $\Pr\left(t_{n-1} \geq \frac{\bar{x} - \mu_0}{s / \sqrt{n}}\right)$

Two - sided P value $2 \times \Pr\left(t_{n-1} \geq \frac{|\bar{x} - \mu_0|}{s / \sqrt{n}}\right)$

28

Stata: One sample t test

- Stata will give you everything you want quite easily
 - “tprob (df, t)”
 - Function returns two-sided P value
 - Performing t test:
 - “ttest var = #val”
 - “ttesti #n #mn #sd #val”
 - provide P values for tests that the mean is equal to #val
 - provide 95% confidence intervals

29

Example: PSA data set

- Test that mean nadir is 15.0 (Why?)

```
. ttest nadir=15
One-sample t test      Number of obs =      50

Varble | Mean StdErr   t   P>|t|   [95% CI]
-----+-----
nadir | 16.36  5.55  2.95  0.005  5.21  27.51
Degrees of freedom: 49

Ho: mean(nadir) = 15
Ha: mean < 15      Ha: mean ~= 15      Ha: mean > 15
t =  0.2450        t =  0.2450        t =  0.2450
P<t=  0.5963      P>|t|= 0.8075      P>t=  0.4037
```

30

Example: Interpretation

- Interpretation
 - The observed results are not atypical of those that might be obtained when the true mean is 15 (P=.81).
 - Based on these results, we cannot with 95% confidence reject the hypothesis that the true mean is 15.
 - We can reject with 95% confidence hypotheses that the true population mean is less than 5.21 or greater than 27.51.

31

Extension to Other Settings

- The one sample tests of means are easily extended to other settings
 - Tests of geometric means
 - Perform inference on log transformed data
 - (Best to “back-transform” by exponentiating)
 - Tests for changes in paired samples
 - Perform inference on differences (or ratios) of measurements within individuals
 - Furthermore, the tests comparing two independent samples look much the same

32