DSA 8010 - Inference on one mean

One numeric variable

One numeric variable

The one sample t procedures are used for inference on one quantitative variable, measured on n individuals. The inferential target is the average value of this variable.

Examples:

- What is the average household income within my school district?
- What is the average age of patients at a clinic?
- How many hours of sleep do middle-schoolers get on average?

If the data are stored in rectangular format, one column of the data set is used for one-sample testing and estimation.

Statistical model for one numeric variable

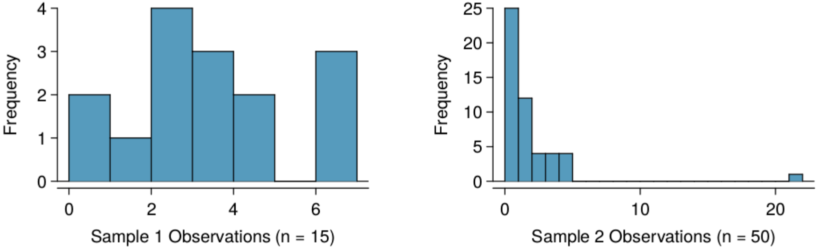
Let $y_i, \quad i = 1, \ldots, n$, be n observed measurements of a numeric variable.

We assume the following statistical model for the one-sample t procedures:
$y_i, \ldots, y_n$ are i.i.d. realizations from a $N(\mu, \sigma^2)$ distribution.

The inferential goal is to learn μ.

Modeling assumptions

Is the statistical model on the previous slide a good approximation for these two datasets?



From *Open Intro Statistics, 4th edition, Diez et al*.

Confidence interval for μ

Sampling distribution of $\bar{y}$

For a one-sample t interval for μ, the point estimate is the sample mean,

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

- The standard error of $\bar{y}$ is $\sigma/\sqrt{n}$. It is estimated from the data using $s/\sqrt{n}$.
- If the data are normal, then the sampling distribution of $\bar{Y}$ is $N\left(\mu, \sigma/\sqrt{n}\right)$.
- If the data are not normal, then in large samples $\bar{Y}$ has an approximately normal distribution.

One-sample t confidence interval for μ

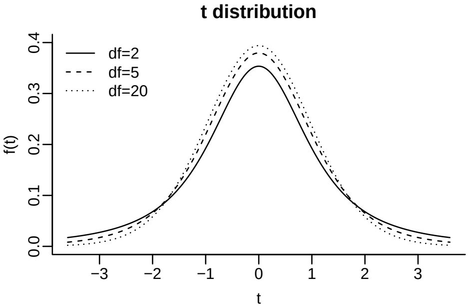
A $(1-\alpha)\cdot 100\%$ confidence interval for μ is

$$\bar{y} \pm t^*_{n-1,\alpha/2} s/\sqrt{n}.$$

- $t^*_{df,\alpha/2}$ refers to the $(1-\alpha/2)\cdot 100$th percentile of the T distribution with degrees of freedom equal to df.
- Find $t^*_{df,\alpha/2}$ using `qt(1-alpha/2,df=n-1)`.
- In small samples (say, $n < 30$), it is important to make sure the data are well-approximated by a Normal distribution. Check for outliers and use a normal quantile plot to check this visually.

The t distribution

The t distribution is a bell-shaped curve with one parameter, the degrees of freedom.



- The t distribution has heavier tails than the normal distribution.
- Intuitively, when estimating μ, we use this distribution instead of the normal distribution to account for our uncertainty in estimating σ using the sample standard deviation.

Example (mercury content)

The FDA's webpage provides some data on mercury content of fish. Based on a sample of 15 croaker white fish (Pacific), a sample mean and standard deviation were computed as 0.287 and 0.069 ppm (parts per million), respectively. The 15 observations ranged from 0.18 to 0.41 ppm. Assume these observations are independent. Make a confidence interval to estimate the mean mercury content of croaker white fish.

From *Open Intro Statistics, 4th edition, Diez et al*.

Example (mercury content)

The FDA's webpage provides some data on mercury content of fish. Based on a sample of 15 croaker white fish (Pacific), a sample mean and standard deviation were computed as 0.287 and 0.069 ppm (parts per million), respectively. The 15 observations ranged from 0.18 to 0.41 ppm. Assume these observations are independent. Make a confidence interval to estimate the mean mercury content of croaker white fish.

From *Open Intro Statistics, 4th edition, Diez et al*.

```
> n<- 15
> tstar <- qt(.995,df=n-1)
> ybar <- 0.287
> s <- 0.069
>
> ybar + tstar*s/sqrt(n)
[1] 0.3400346
> ybar - tstar*s/sqrt(n)
[1] 0.2339654
```

We are 99% confident the the mean mercury content of croaker white fish is between 0.2340 ppm and 0.3400 ppm.

Interpretation of confidence intervals

- Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. They found the following 95% confidence interval for the mean lead concentration in the blood to be ($112\mu g/l$, $142\mu g/l$).

  We are 95% confident that the mean lead concentration in the blood of police officers working traffic enforcement is between $112\mu g/l$ and $142\mu g/l$.

Interpretation of confidence intervals

- Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. They found the following 95% confidence interval for the mean lead concentration in the blood to be $(112\mu g/l, 142\mu g/l)$.

  Not a correct interpretation: There is a 95% chance that a randomly-selected police officer working in traffic enforcement will have lead concentration in the blood between $112\mu g/l$ and $142\mu g/l$.

Interpretation of confidence intervals

- Government officials in a particular county are concerned lead exposure among the 64 police officers who working in traffic enforcement in a particular precinct. They took blood samples from each of the 64 officers and found the following 95% confidence interval for the mean lead concentration in the blood among officers in the precinct to be (112$\mu g/l$, 142$\mu g/l$).

  If data are collected from an entire population of interest, a confidence interval is not particularly meaningful.

Sample size and estimation

A theater instructor wants to stock a costume closet for students and wants an idea of average height of teenage boys. He collects a random sample of 8 teenagers and finds that the sample average was 63 inches with a standard deviation of 5.9 inches.

Find the margin of error for a 95% confidence interval for μ.

Sample size and estimation

Now suppose he collects a random sample of 80 teenagers. How will the margin of error change? (Assume the sample standard deviation is unchanged.)

Sample size and estimation

Suppose he collects a random sample of 80 teenagers, but now wants to make a 90% confidence interval. How will the margin of error change? (Assume the sample standard deviation is unchanged.)

Sample size formula for inference on one mean

Margin of error of a confidence interval for μ:

$$t^*_{n-1,\alpha/2} s/\sqrt{n}$$

.

- All other things equal, higher confidence levels produce higher margins of error (wider CIs).
- All other things equal, larger sample sizes produce smaller margins of error (narrower CIs).

Increasing sample size will increase the precision of an estimate.

Hypothesis test for μ

Hypothesis test for μ.

The one-sample t test:

Hypotheses. Null hypothesis: $H_0 : \mu = \mu_0$;
Alternative hypothesis: $H_a : \mu \neq \mu_0$ (or $<$, $>$).

Test statistic.

$$t_0 = \frac{(\bar{y} - \mu_0)}{s/\sqrt{n}}.$$

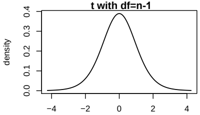
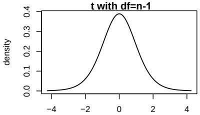
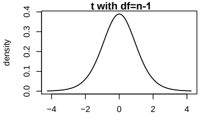
An estimate of how many standard errors $\bar{y}$ is away from μ_0.

P-value. Two-sided alternative: $2 * P(T > |t_0|)$ where t has a T distribution with n-1 degrees of freedom.
`2*pt(abs(t0),df=n-1,lower.tail=FALSE).`

Decision. Reject H_0 if the p-value is less than α.

One-sided alternatives

The p-value is the T distribution tail probability "in the direction of the alternative."

Type	Alternative	p-value	p-value sketch
Two-sided	$H_A : \mu \neq \mu_0$	$2 \cdot P(T > \lvert t_0 \rvert)$	
Right-sided	$H_A : \mu > \mu_0$	$P(T > t_0)$	
Left-sided	$H_A : \mu < \mu_0$	$P(T < t_0)$	

Example (race times)

The average time for all runners who finished the Cherry Blossom Race, a 10-mile race in Washington, DC, in 2006 was 93.29 minutes (93 minutes and about 17 seconds). Data is collected from a random sample of 100 participants in the 2017 Cherry Blossom Race. The sample mean and sample standard deviation of the sample of 100 runners from the 2017 Cherry Blossom Race are 97.32 and 16.98 minutes, respectively. Do the data provide evidence that the runners have, on average, slowed since 2006?

From *Open Intro Statistics, 4th edition, Diez et al*

Example (race times)

The average time for all runners who finished the Cherry Blossom Race, a 10-mile race in Washington, DC, in 2006 was 93.29 minutes (93 minutes and about 17 seconds). Data is collected from a random sample of 100 participants in the 2017 Cherry Blossom Race. The sample mean and sample standard deviation of the sample of 100 runners from the 2017 Cherry Blossom Race are 97.32 and 16.98 minutes, respectively. Do the data provide evidence that the runners have, on average, slowed since 2006?

From *Open Intro Statistics, 4th edition, Diez et al*

```
> ybar <- 97.32
> mu0 <- 93.29
> s <- 16.98
> n <- 100
>
> # test statistic
> t0 <- (ybar - mu0)/(s/sqrt(n))
> t0
[1] 2.37338
>
> # p-value
> pt(t0,df=n-1,lower.tail=FALSE)
[1] 0.009778332
```

Example (race times)

Make a 95% confidence interval for the mean race time of runners in 2017.

```
> # CI for mu
> ybar <- 97.32
> s <- 16.98
> n <- 100
> alpha <- 0.05
> tstar <- qt(1-alpha/2, df=n-1)
>
> ybar -tstar*s/sqrt(n)
[1] 93.9508
> ybar + tstar*s/sqrt(n)
[1] 100.6892
```

Confidence interval as an inversion of a two-sided hypothesis test

We sometimes say that a confidence interval is an "inversion" of a hypothesis test.

- If μ_0 is contained in the $(1-\alpha) \times 100\%$ CI for μ, then we would fail to reject the null $H_0 : \mu = \mu_0$ against the two-sided alternative, $H_A : \mu \neq \mu_0$ using an α threshold to reject H_0. If μ_0 is not contained in the CI, then we would reject H_0.

Power and type 2 error

Type 1 and type 2 errors

A type 1 error occurs when H_0 is rejected, but it is true.

A type 2 error occurs when H_0 is false, but it is not rejected.

- The α value for a hypothesis test is a fixed value that gives the probability of making a Type 1 error.
- The probability of type 2 error is not fixed when a hypothesis test is performed. The larger the sample is, the less likely a type 2 error is.

Power and type 2 error

Type 1 and type 2 errors

A type 1 error occurs when H_0 is rejected, but it is true.

A type 2 error occurs when H_0 is false, but it is not rejected.

- The α value for a hypothesis test is a fixed value that gives the probability of making a Type 1 error.
- The probability of type 2 error is not fixed when a hypothesis test is performed. The larger the sample is, the less likely a type 2 error is.
- Power is defined as the probability of rejecting H_0 when H_0 is false, or 1-P(Type 2 error).
- Power increases with increasing sample size, all other things equal.
- There are formulas for calculating required sample size to achieve a certain power.