General form of a confidence interval

In general, a CI for a parameter has the form

$$\text{estimate} \pm \text{margin of error}$$

where the margin of error is determined by the confidence level $(1 - \alpha)$, the population SD σ, and the sample size N.

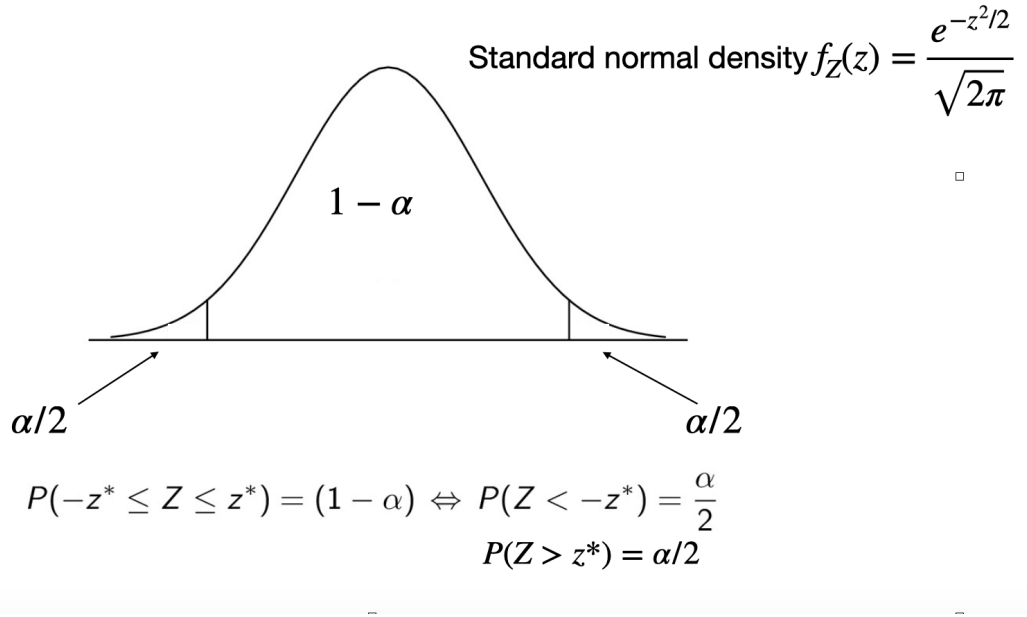
A $(1 - \alpha)$ confidence interval for a parameter θ is an interval computed from a SRS by a method with probability $(1 - \alpha)$ of containing the true θ.

For a random sample of size N drawn from a population of unknown mean μ and known SD σ, a $(1 - \alpha)$ CI for μ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{N}}$$

Finding the critical value z^*

z^* is the **critical value**, selected so that a standard Normal density has area $(1-\alpha)$ between $-z^*$ and z^*.



The quantity $z^*\sigma/\sqrt{N}$, then, is the **margin error**.

If the population distribution is normal, the interval is *exact*. Otherwise, it is *approximately correct for large N*.

Tests for population mean

Suppose we want to test the hypothesis that μ has a specific value:

$$H_0 : \mu = \mu_0$$

Since $\bar{x}$ estimates μ, the test is based on $\bar{x}$, which has a (perhaps approximately) Normal distribution. Thus,

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N}}$$

is a standard normal random variable, *under the null hypothesis*.
p-value for $H_a : \mu \neq \mu_0$ is $2P(Z \geq |z|)$ (area of both tails)

To reject at p-value $= \alpha$, you need: $|z| \geq z^* = z_\alpha$

Test for population proportion I

When population is just 0/1's population mean is simply proportion of 1's. One option is to use the same kind of test as above. (We will get back to that)

However in this case, once you make a Null hypothesis on the population proportion $H_0 : p = p_0$, you have fully determined the distribution of the number of 1's in the sample: $B(N, p_0)$.

So you can compute $\Delta = |\sum_{n=1}^{N} x_n - Np_0|$ and get the p-value: $P(S > Np_0 + \Delta) + P(S < Np_0 - \Delta)$ for $S \sim B(N, p_0)$.

Test for population proportion II

Example: You want to test Trump's hypothesis that there were many illegitimate voters who voted against him in 2016. I.e. he should have larger percent of the votes.
Your Null hypthesis is that that the proportion of Trump supporters was indeed $p = p_0 = .461$, against the alternative that $p \neq p_0$
You take a random sample of 1000 people, you keep only those who voted and were legitimate registered voters, say 534 are left. You find that 253 voted for Trump. Under the null we expect $P = p_0 * 534 = 246$.

```
P=.461*534
Delta=abs(P-253)
Pv=pbinom(P-Delta,534,.461)+(1-pbinom(P+Delta,534,.461))
Pv
```

```
[1] 0.5436
```

Test for population proportion III

So the P-value is .543! If the Null is true there is 54% chance of getting more than 253=246+(253-246) or less than 246-(253-246)=239.
You therefore reject Trumps Hypothesis...

Test Interpretations

Saying that a result is *statistically significant* does not signify that it is large or necessarily important. That decision depends on the particulars of the problem. A statistically significant result only says that there is substantial evidence that H_0 is false.
Failure to reject H_0 does not imply that H_0 is correct. It only implies that *we have insufficient evidence to conclude that* H_0 *is incorrect*.

Hypothesis tests and CI's I

A level α two-sided test rejects a hypothesis $H_0 : \mu = \mu_0$ exactly when the value of μ_0 falls outside a $(1 - \alpha)$ confidence interval for μ.

For example, consider a two-sided test of the following hypotheses

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

at the significance level $\alpha = .05$.

Assume the test statistic is z and $2P(Z > |z|) = 2P(Z > z) = p < \alpha$. Let z_α be the critical value for level α. Assume the population SD is σ_0.

Hypothesis tests and CI's II

$$
\begin{aligned}
& p < \alpha \\
& \Updownarrow \\
& z > z_\alpha \quad \text{or} \quad z < -z_\alpha \\
& \Updownarrow \\
& \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{N}} > z_\alpha \quad \text{or} \quad \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{N}} < -z_\alpha \\
& \Updownarrow \\
& \mu_0 < \bar{x} - z_\alpha \cdot \frac{\sigma_0}{\sqrt{N}} \quad \text{or} \quad \mu_0 > \bar{x} + z_\alpha \cdot \frac{\sigma_0}{\sqrt{N}} \\
& \Updownarrow \\
& \mu_0 \notin [\bar{x} - z_\alpha \cdot \frac{\sigma_0}{\sqrt{N}}, \bar{x} + z_\alpha \cdot \frac{\sigma_0}{\sqrt{N}}]
\end{aligned}
$$

μ_0 is not in the α confidence interval if and only if the null hypothesis is rejected at the α level.

Hypothesis tests and CI's III

- If μ_0 is a value inside the 95% confidence interval for μ, then this test will have a p-value greater than .05, and therefore will not reject H_0.
- If μ_0 is a value outside the 95% confidence interval for μ, then this test will have a p-value smaller than .05, and therefore will reject H_0.

Inference for the mean I

We have already seen that when we take a SRS, $X_1, X_2, \ldots, X_N$, from a population with unknown μ and known σ, if either

- the population is Normally distributed, or
- N is large enough,

then

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{N}}\right)$$

Equivalently,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \sim N(0, 1)$$

What if the population SD σ is unknown? (A far more likely occurrence.)

Inference for the mean II

If the true population SD, σ, is **unknown**, we estimate σ using the *sample* standard deviation denoted $S = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(X_N - \bar{X})^2}$.

S^2 is an unbiased estimate of σ^2, but is a random quantity.

Now, instead of dealing with

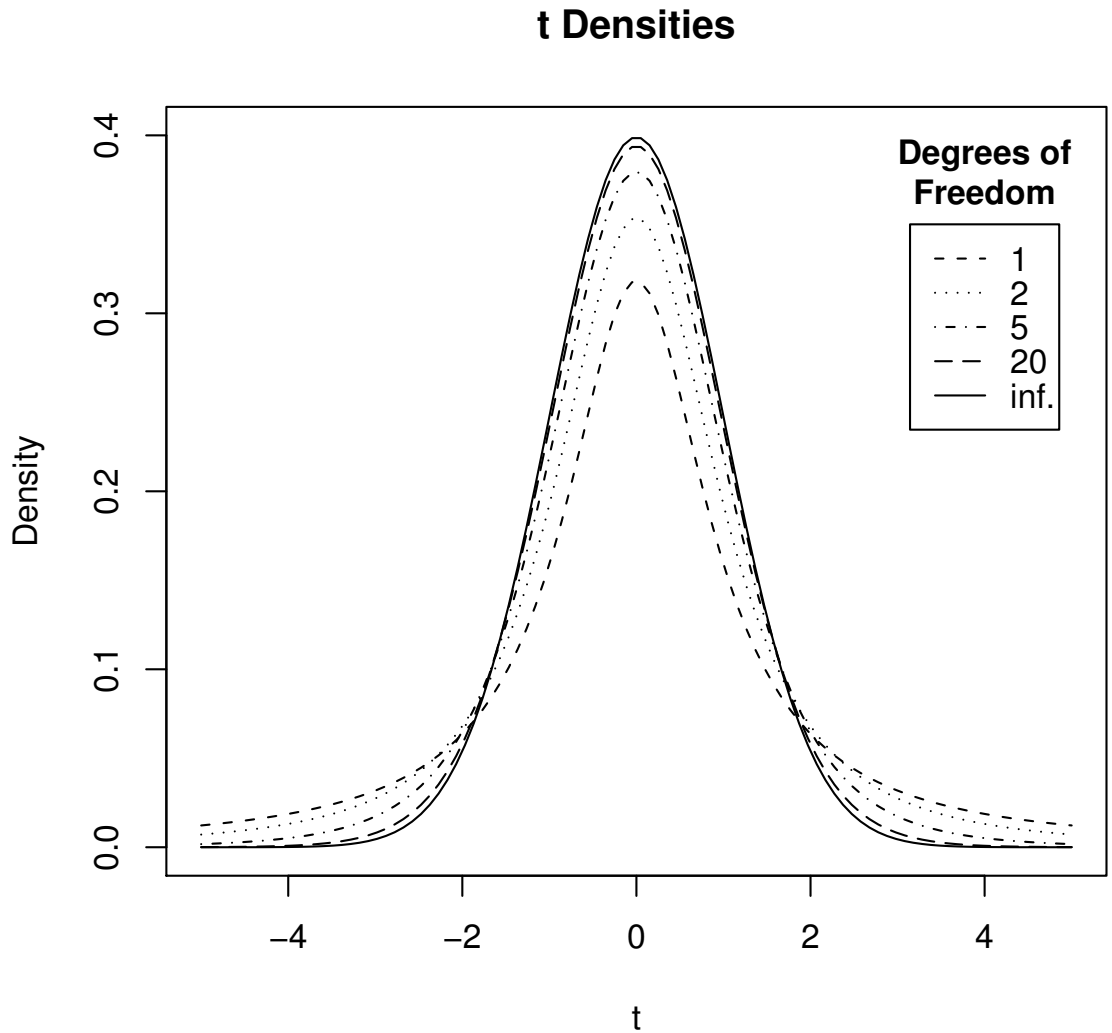
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \sim N(0,1)$$

we are interested in the quantity

$$T = \frac{\bar{X} - \mu}{S/\sqrt{N}} \sim t_{(N-1)}$$

Here, $t_{(N-1)}$ is **Student's** t **distribution**, with $n-1$ *degrees of freedom*.

Inference for the mean III



Inference for the mean IV

(Unlike the Normal or Binomial distributions, each of which has two parameters, the t distribution has only one parameter, called the **degrees of freedom**.)

Properties of the t Distribution

- Symmetric about zero
- Bell-shaped – similar to normal distribution
- More spread out than normal – heavier tails
- Exact shape depends on the degrees of freedom
- As the number of degrees of freedom increases, the t distribution converges to the Normal distribution.

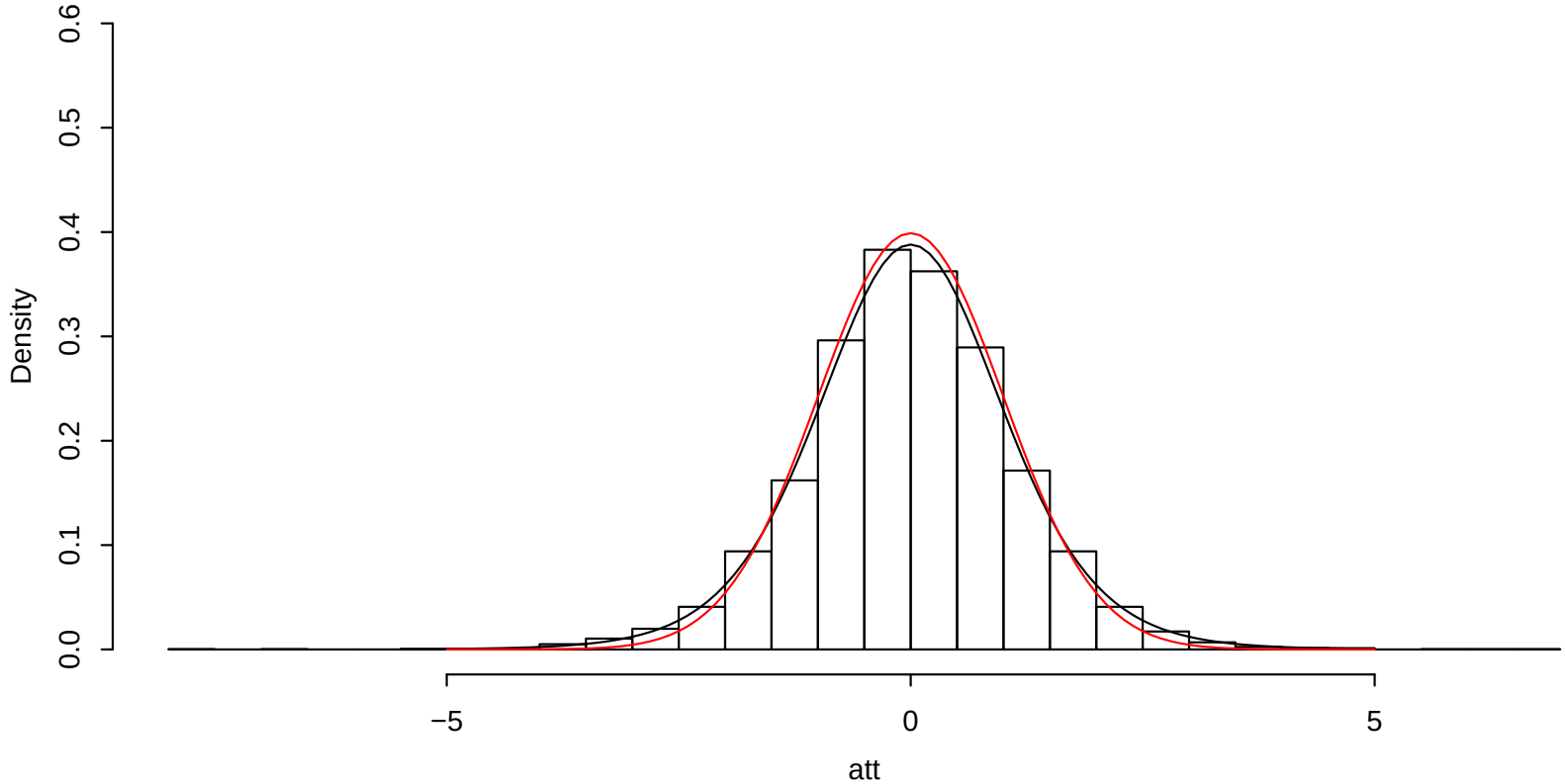
Inference for the mean V

Generate 10000 random samples of size 10 from a $N(0, 1)$ distribution and look at the histogram of 10000- T statistics.

```
aa=replicate(10000,rnorm(10))
ave=apply(aa,2,mean)
sds=apply(aa,2,sd)
att=(ave/(sds/sqrt(10)))
```

```
hist(att,breaks=50,freq=F,main="",ylim=c(0,.6))
x=seq(-5,5,.1)
y=dt(x,9)
lines(x,y)
yn=dnorm(x)
lines(x,yn,col="red")
```

Inference for the mean VI



Overlayed in black is T_9 vs. the $N(0,1)$ in red. The black curve fits histogram better, especially in the tails.

Inference for the mean VII

Confidence Intervals with Unknown σ

Recall that, for a population with unknown μ and *known* σ, $100(1-\alpha)\%$ CI for μ, based on an SRS $X_1, X_2, \ldots, X_N$ is given by $\left(\bar{X} - z_\alpha \frac{\sigma}{\sqrt{N}},\ \bar{X} + z_\alpha \frac{\sigma}{\sqrt{N}}\right)$.

If σ is unknown,

$$\frac{\bar{X} - \mu}{S/\sqrt{N}} \sim t_{(n-1)}$$

so we substitute a t-critical value, t^* for z_α:

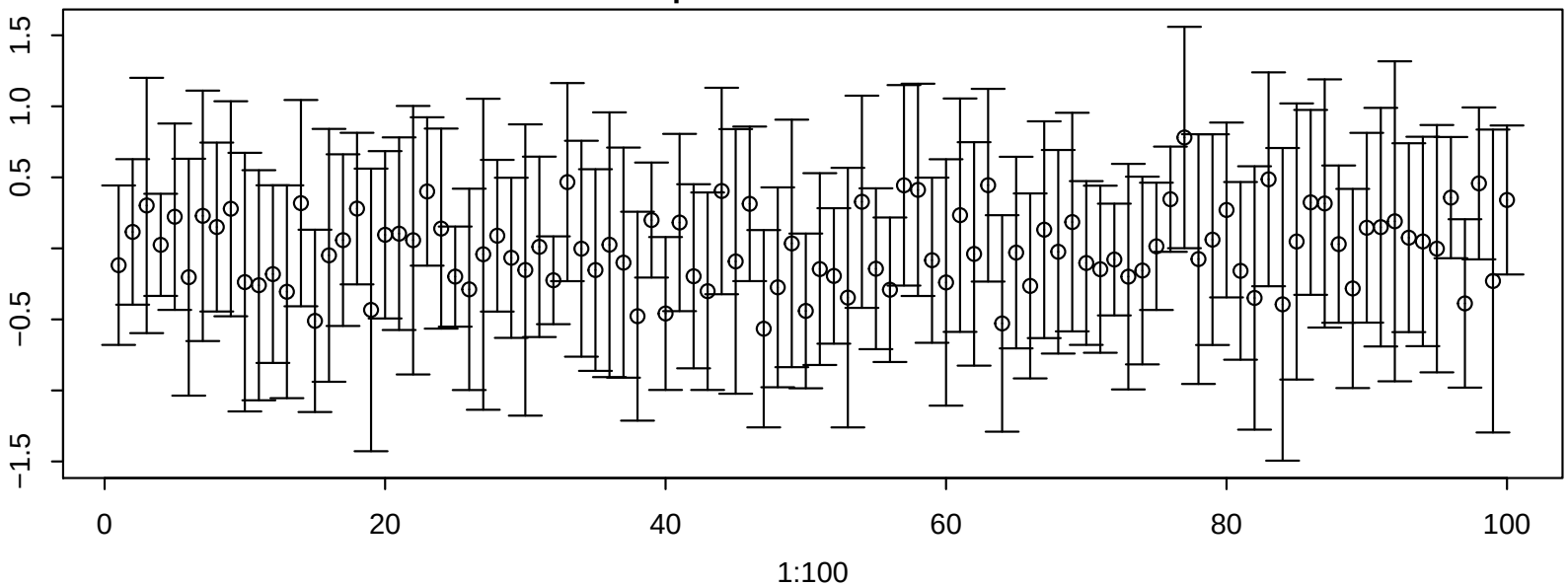
$$\left(\bar{X} - t^* \frac{s}{\sqrt{N}},\ \bar{X} + t^* \frac{s}{\sqrt{N}}\right)$$

The critical value, $t^* = t_{N-1,\alpha}$, is chosen such that $100(1-\alpha)\%$ of the area under the $t_{(N-1)}$ density lies between $-t^*$ and t^*.

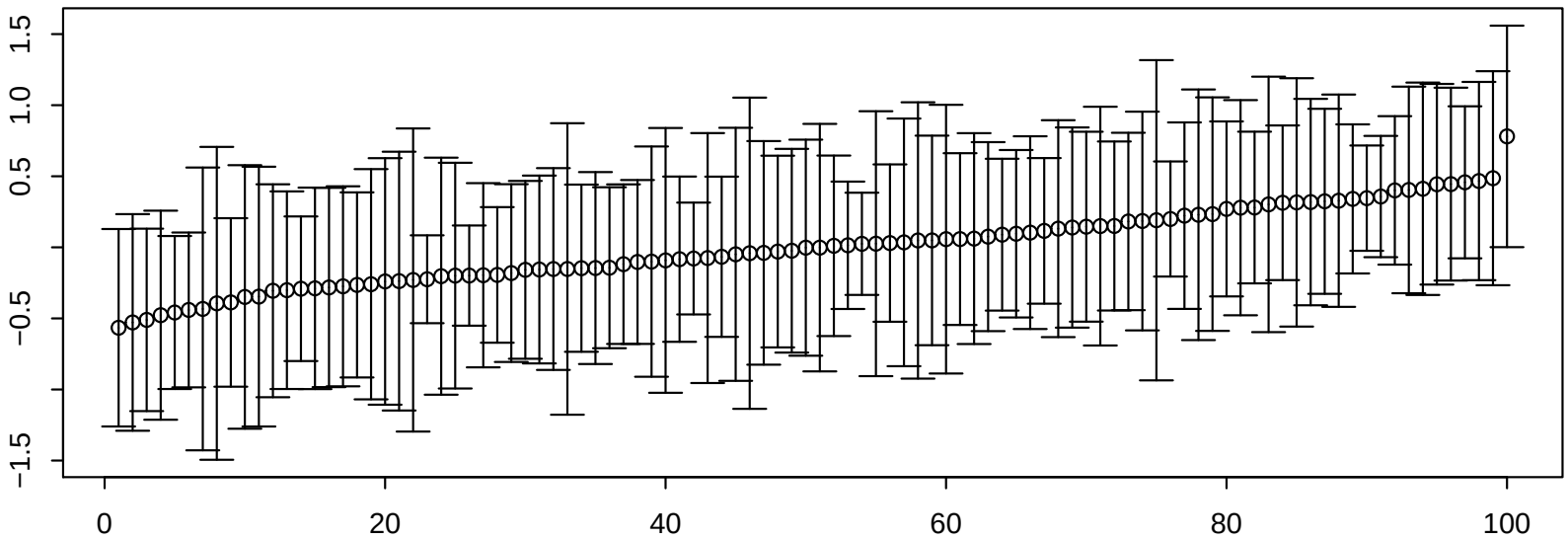
These confidence intervals are again random. In addition to having a random center $\bar{X}$ they have a random *width* $t^* S/\sqrt{N}$.

Inference for the mean VIII

We show 100 CI's for size 10 samples from the standard normal:



Inference for the mean IX



In the second plot they are plotted according to increasing order of the means.

Inference for the mean X

We find the fraction of intervals that don't include the true mean - very close to .05.

```
mean(ave[1:100]+qt(.975,9)*sds[1:100]/sqrt(10) < 0)+
mean(ave[1:100]-qt(.975,9)*sds[1:100]/sqrt(10) > 0)
```

```
[1] 0.01
```

Inference for the mean XI

(1) If the underlying population is Normally distributed, the interval is exact.
(2) Otherwise, the interval is approximately correct if N is not too small (say, $N \geq 15$), the data are not strongly skewed, and there are no outliers.
(3) With N sufficiently large (say $n \geq 40$), the approximation is correct even if the data are clearly skewed.

One-sample t-test I

Suppose a SRS of size N is drawn from a $N(\mu, \sigma)$ population with both μ and σ unknown. The t-statistic,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{N}}$$

has the t distribution with $N - 1$ d.f.
To test $H_0 : \mu = \mu_0$, compute the one-sample t statistic,

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{N}}$$

The p-value for the two sided test is $2P(T_{N-1} \geq |t|)$.
The critical value is $t^* : P(|T_{N-1}| > t^*) = 2P(T_{N-1} > t^*) = \alpha$.
Check if $|t| \geq t^*$

One-sample t-test II

For a one-sided alternative. $H_a : \mu > \mu_0$, p-value is $P(T_{N-1} \geq t)$.
Critical value is $t^* : P(T_{N-1} > t^*) = \alpha$
For a one-sided alternative. $H_a : \mu < \mu_0$, p-value is $P(T_{N-1} < t)$.
Critical value is $t^* : P(T_{N-1} < -t^*) = \alpha$
This is exact if the population is normal, and otherwise approximately correct for large N.

One-sample t-test III

Example: Growth of Tumor

Let X (in mm) denote the growth in 15 days of a tumor induced in a mouse. It is known from a previous experiment that the average tumor growth is $4mm$.

A sample of 20 mice that have a genetic variant hypothesized to be involved in tumor growth yielded $\bar{x} = 3.8mm, s = 0.3mm$.

Test whether $\mu = 4$ or not, assuming growths are normally distributed.

1. State the hypotheses

$$H_0 : \mu = 4 \qquad H_a : \mu \neq 4$$

One-sample t-test IV

2. Calculate the t-statistic

$$t = \frac{3.8 - 4.0}{0.3/\sqrt{20}} = -2.98$$

3. Determine the p-value

$$p = 2\mathrm{P}(T_{19} \geq 2.98) = 0.008$$

Since p is less than 0.01, we reject H_0 at significance level $\alpha = 0.01$. There is evidence that the population mean growth is not $4mm$.

One-sample t-test V

What if we wanted a 99% CI for μ instead? The CI is given by

$$\left(\bar{x} - t^* \frac{s}{\sqrt{N}},\ \bar{x} + t^* \frac{s}{\sqrt{N}}\right)$$
$$= \left(3.8 - 2.861 \times \frac{0.3}{\sqrt{20}},\ 3.8 + 2.861 \times \frac{0.3}{\sqrt{20}}\right)$$
$$= (3.61,\ 3.99)$$

where t^* : $P(|T_{19}| > .t^*) = .01 \rightarrow t^* = 2.861$
Note that 4 is outside this CI. From this, we can draw the same conclusion as from the test. Namely, at significance level $\alpha = 0.01$, the mean growth not equal to $4mm$.

One-sample t-test VI

Confidence Intervals and Two-Sided Tests

A two-sided hypothesis test with significance level α rejects the null hypothesis $H_0 : \mu = \mu_0$ if and only if the value μ_0 falls outside the $100(1-\alpha)\%$ CI for μ.
Reporting a CI is generally more informative than just reporting a p-value or the decision made on the basis of a hypothesis test.

One-sample t-test VII

One-Sided Alternatives

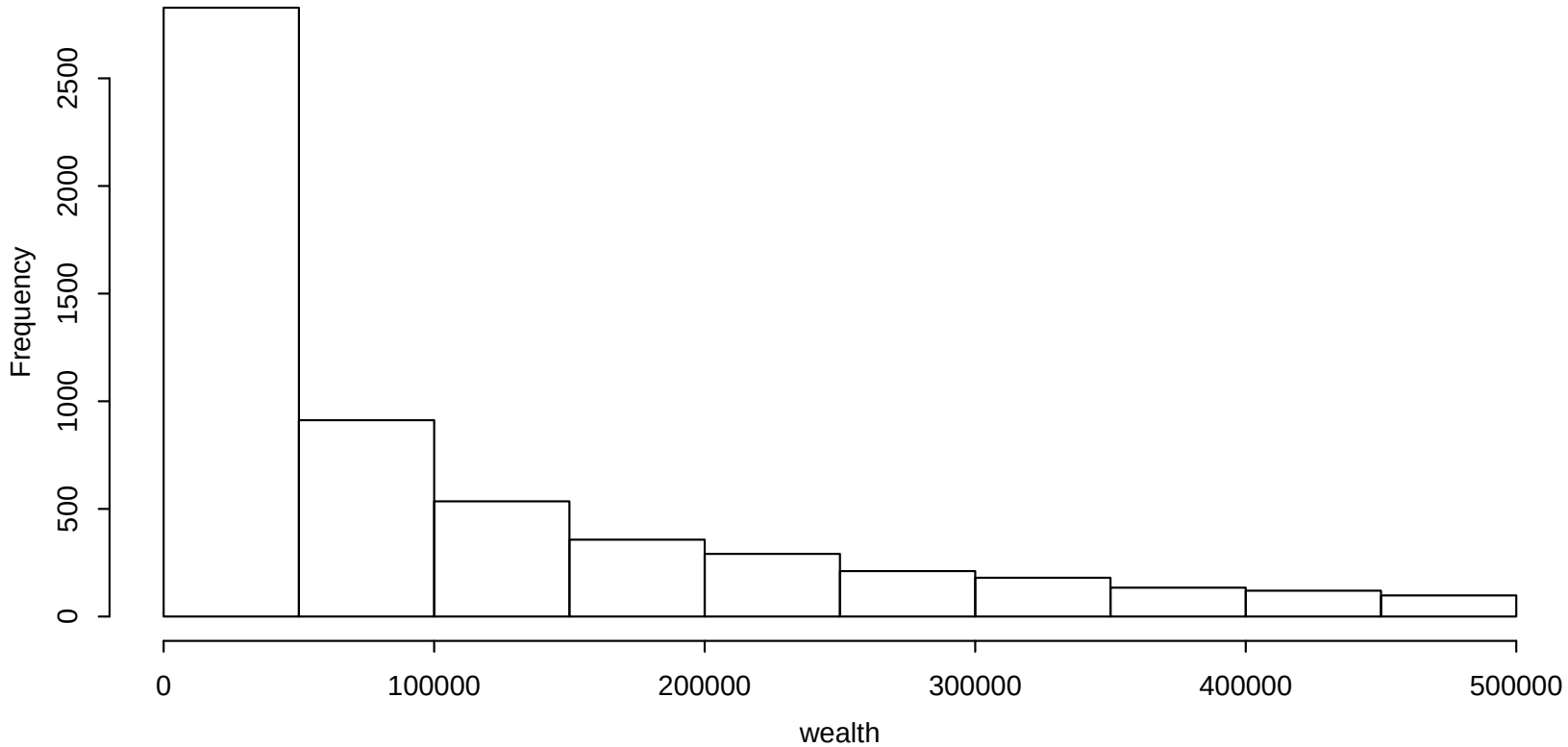
In the previous example, suppose we wished to test whether $\mu < 4$.

1. State the hypotheses $H_0 : \mu = 4 \qquad H_a : \mu < 4$
2. Calculate the t-statistic $t = \frac{3.8-4}{0.3/\sqrt{20}} = -2.98$
3. Determine the p-value $p = P(T_{19} \leq -2.98)$

Since $\mathrm{P}(T_{19} \geq 2.861) = 0.005$, p is less than 0.005. Thus, we reject H_0 at significance level $\alpha = 0.01$. There is evidence that the population mean growth is less than 4 millimeters.

Sometimes T-dist doesn't fit I

Histogram of population wealth between 0 and 500000:

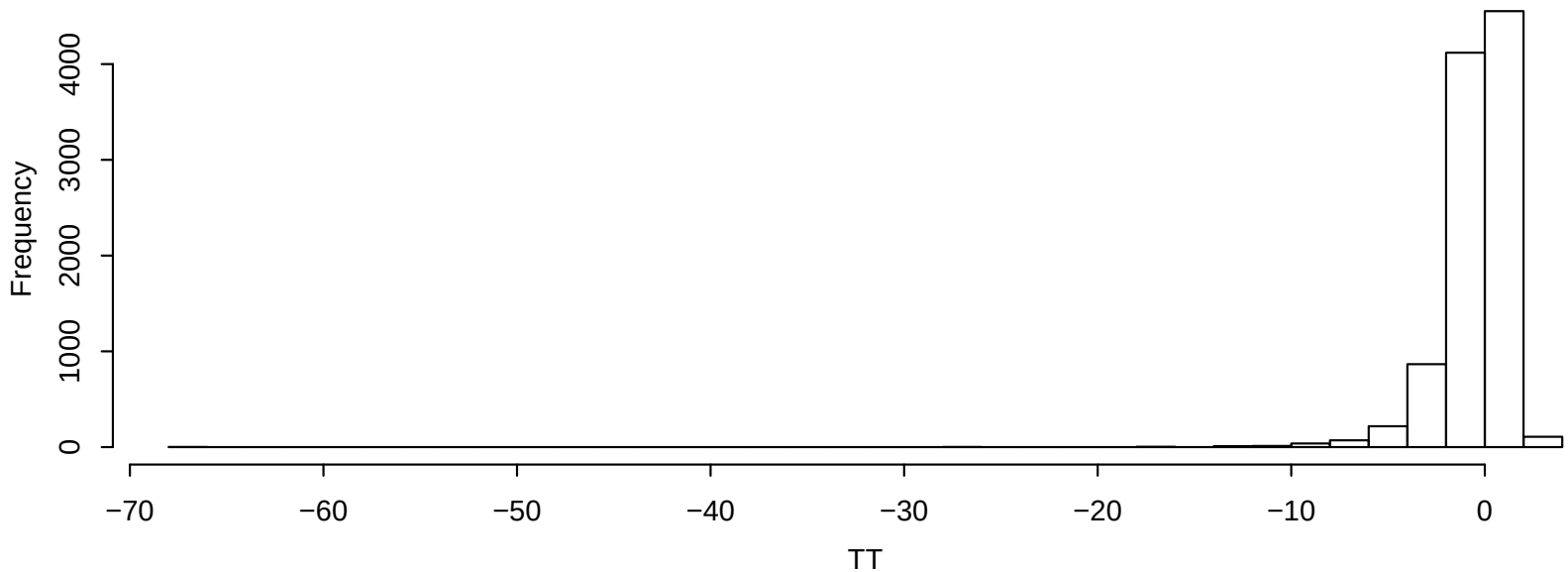


Mean is about 100000

Sometimes T-dist doesn't fit II

Histogram of T test-statistic for samples of size 10.

```
sampsize=10
wealths=replicate(10000,sample(wealth,sampsize))
ave=apply(wealths,2,mean);sds=apply(wealths,2,sd);
TT=(ave-MEAN)/(sds/sqrt(sampsize))
hist(TT,breaks=40,main="")
```



Histogram is very skewed.

Sometimes T-dist doesn't fit III

Imagine our sample was:

```
Y

 [1] 189500  55600  14800  22000   1000  48600  16600  56000
 [9]  19000 102105

c(mean(Y),sd(Y))

[1] 52520 56452

t=(mean(Y)-100000)/(sd(Y)/sqrt(10))
PY=2*(1-pt(abs(t),9))
c(t,PY)

[1] -2.65966  0.02606
```

Sometimes T-dist doesn't fit IV

We want to test the hypothesis that mean wealth is 100000 vs. the alternative that it is not.

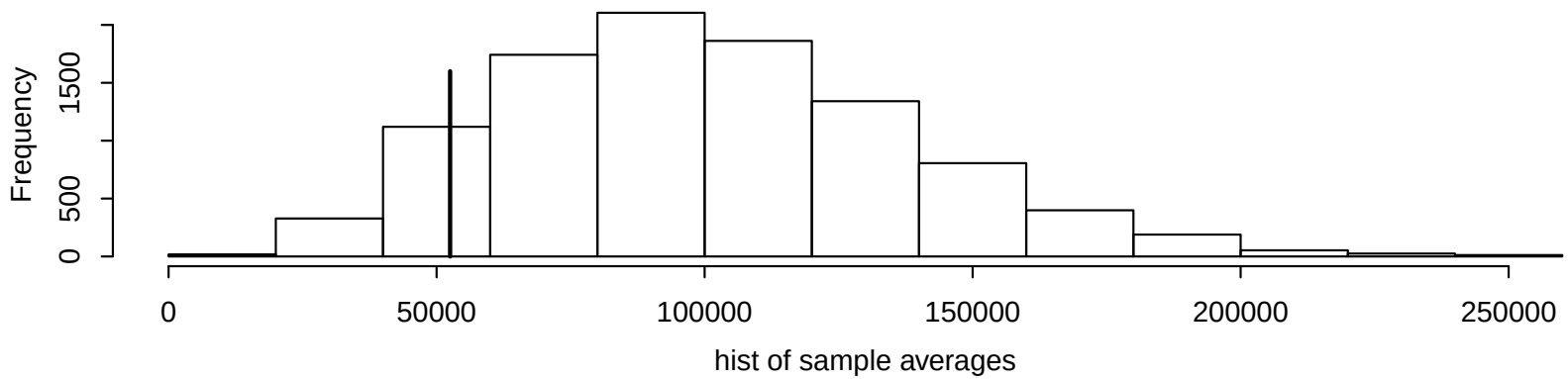
According to the T-test the p-value is 0.0261 - quite small.

Sometimes T-dist doesn't fit V

Since in this case we know the full population we can check what proportion of the 10000 samples of size 10 from the population have a mean that is further than 52520.5 from 100000 (above or below).

```
del=100000-my
mean(ave<100000-del | ave > 100000+del)
```

```
[1] 0.2046
```



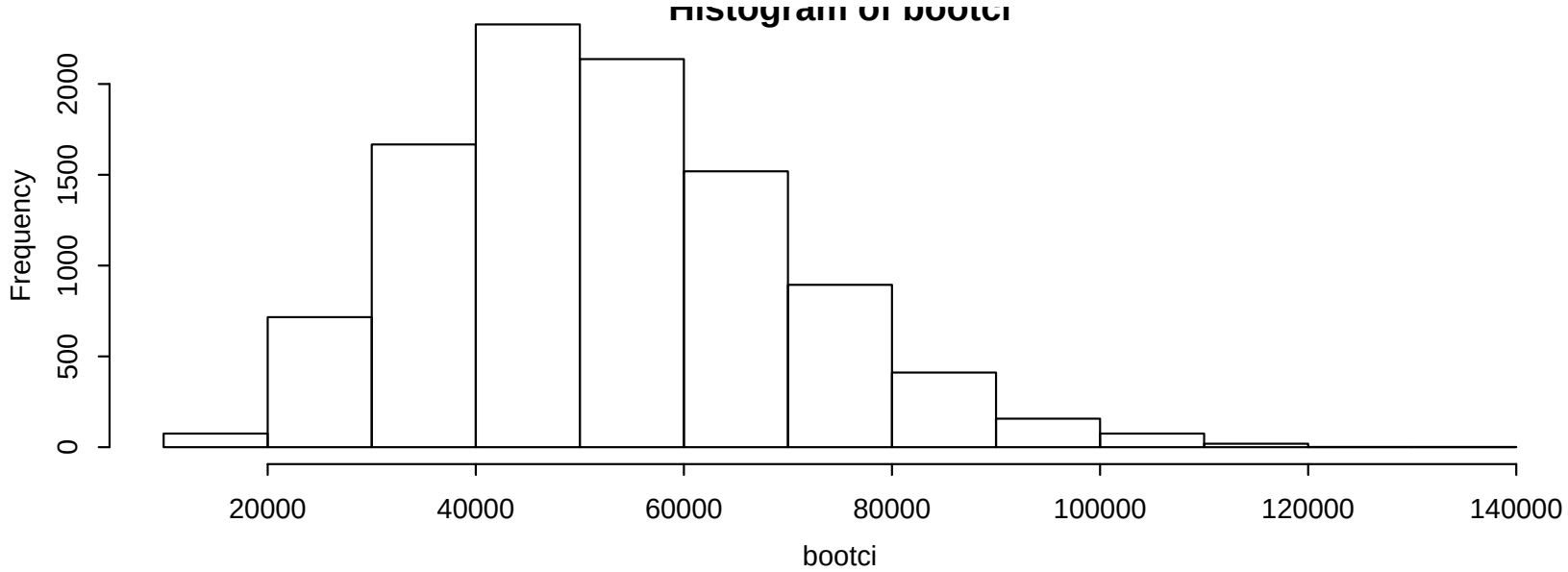
So the mean of 52520.5 is **not** so unlikely at all under the null.

Bootstrap I

In real life we only have **one** sample to work with, we can't sample multiple times from the population.
So we pretend Y describes the population. We can draw 10000 times from the sample Y and look at the distribution of means estimated from each of these samples, and find the $\alpha/2$ and $1-\alpha/2$ quantiles. If the histogram is skewed (as it is below) its an indication that the T-test may not be valid.

Bootstrap II

```
bootci=replicate(10000,mean(sample(Y,10,replace=TRUE)))
hist(bootci)
```




```
quantile(bootci,c(.025,.975))
```

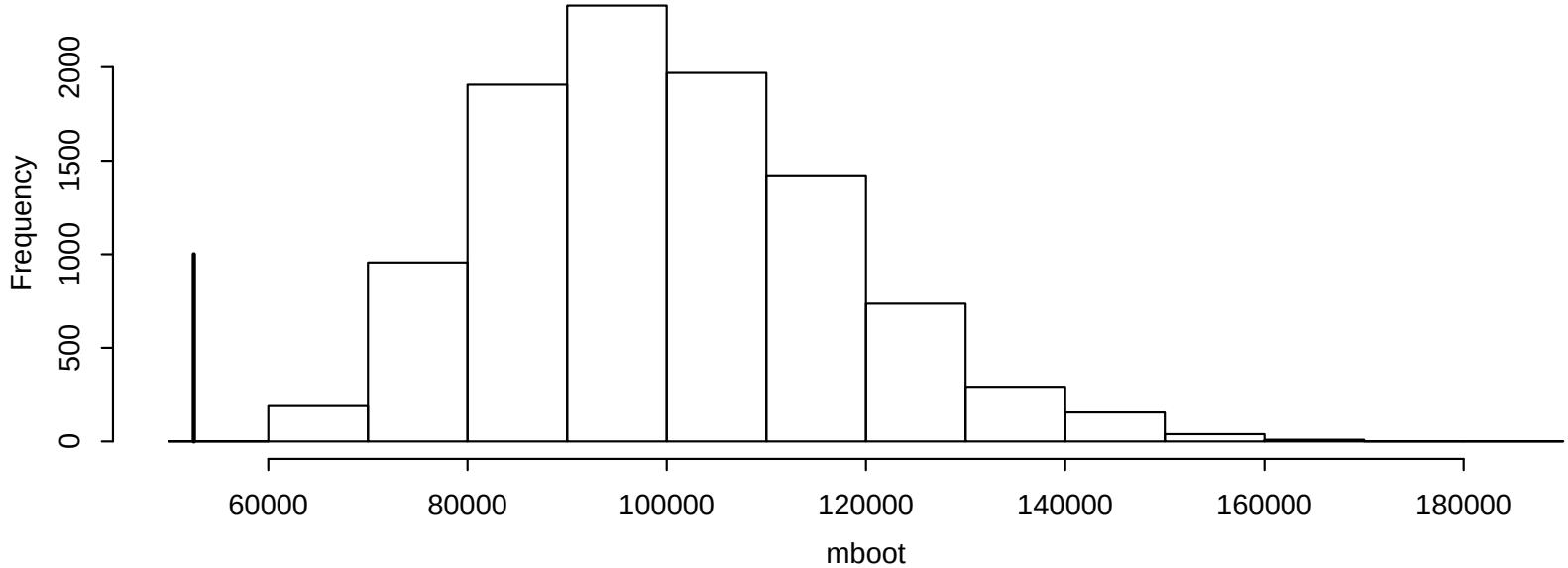
```
 2.5% 97.5%
23970 90191
```

Bootstrap III

For hypothesis testing shift sample to center iit around the hypothesized mean. Then draw 10000 samples with replacement, compute the mean and look at the histogram.

```
my=mean(Y)
YH=Y-my+100000
boot=replicate(10000,sample(YH,10,replace=TRUE))
mboot=apply(boot,2,mean)
hist(mboot,main="")
lines(c(my,my),c(0,1000),lwd=2)
```

Bootstrap IV



We can use the multiple computed means to test the original statistic directly:

```
del=100000-my
mean(mboot<100000-del | mboot>100000+del)
```

```
[1] 0.0074
```

Bootstrap V

So according to the bootstrap sample the observed mean is very unlikely under the null...

The main conclusion should be that because the histogram of means is skewed **something is wrong** and there is no easy fix.

Inference for proportion I

Suppose we want to estimate the proportion p of some characteristic of a population, and we undertake the following procedure:

1. Draw a SRS of size N.
2. Record the number X of "successes" (those individuals having the characteristic).
3. Estimate the unknown true population proportion p with the sample proportion of successes $\hat{p} = \frac{X}{N}$

p is the mean of the population, but in this case it determines the $SD = p(1-p)$. In the general case σ is not determined by μ.
What is the sampling distribution of $\hat{p}$?

Inference for proportion II

If n is sufficiently large – i.e. if

$$np \geq 10 \text{ and } N(1-p) \geq 10,$$

then

$$\hat{p} \dot{\sim} N\left(p, \sqrt{\frac{p(1-p)}{N}}\right)$$

Thus, an approximate $(1-\alpha)$ CI for the population proportion p is given by

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

where z^* is chosen so that $P(Z > z^*) = \alpha/2$ for $Z \sim N(0,1)$.

Inference for proportion III

What if we want to test whether $p = p_0$ for some fixed value p_0?
The null hypothesis is $p = p_0$, and under this hypothesis,

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \dot{\sim} N(0,1)$$

Notice that we are using a different value for the SD of $\hat{p}$ than was used for the CI. Since H_0 specifies a true value for p, the SD of $\hat{p}$ under H_0 is given by $\sqrt{\frac{p_0(1-p_0)}{N}}$
The p-values for this test are:

- $H_a : p > p_0 \qquad P(Z \geq z)$
- $H_a : p < p_0 \qquad P(Z \leq z)$
- $H_a : p \neq p_0 \qquad 2P(Z \geq |z|)$

for $Z \sim N(0,1)$.
Some care needs to be taken when p is very close to 0 or 1.

Inference for proportion IV

A random sample of 2700 California lawyers revealed only 1107 who felt that the ethical standards of most lawyers are high (*AP, Nov. 12, 1994*).

1. Does this provide strong evidence for concluding that fewer than 50% of all California lawyers feel this way?
2. What is a 90% confidence interval for the true proportion of California lawyers who feel that ethical standards are high?

Inference for proportion V

```
hp=1107/2700
Z=(hp-.5)/sqrt(.25/2700)
Z^2

[1] 87.48

prop.test(1107,2700,conf.level = .90,correct=FALSE)


1-sample proportions test without continuity
correction

data:  1107 out of 2700
X-squared = 87, df = 1, p-value <2e-16
alternative hypothesis: true p is not equal to 0.5
90 percent confidence interval:
 0.3945 0.4257
sample estimates:
   p
0.41
```