STAT 113
Standardized Statistics

Colin Reimer Dawson

Oberlin College

November 3, 2017

Outline

Standard Normal

CIs from a Standard Normal

P-values Using a Standard Normal

Goals

Confidence Intervals

If we can replace the bootstrap distribution with a Normal model, we can construct a confidence interval.

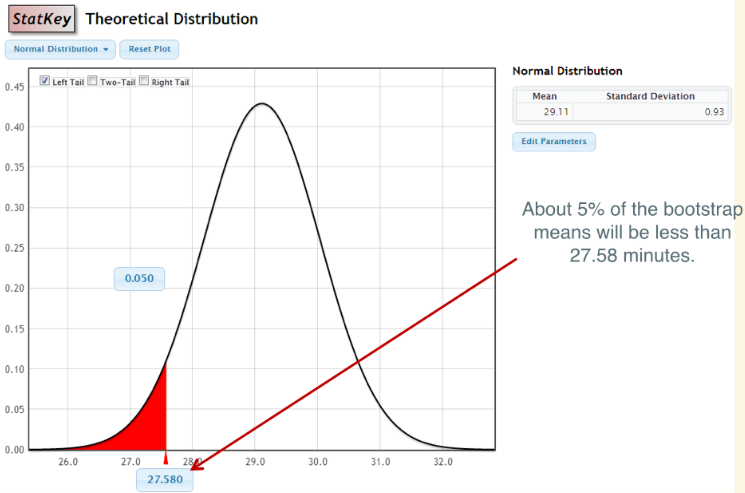
P-values

If we can replace a randomization distribution with a Normal model, we can compute P-values.

Quantiles of a Normal Curve

Suppose that the bootstrap distribution of means for samples of size 500 Atlanta commute times is $\mathcal{N}(29.11, 0.93)$. Find an endpoint (percentile) so that just 5% of the bootstrap means are smaller.

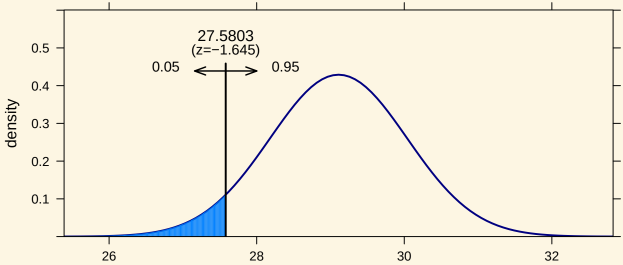
StatKey...


StatKey
Theoretical Distribution
Normal Distribution
Reset Plot
Left Tail
Two-Tail
Right Tail
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.050
26.0
27.0
28.0
29.0
30.0
31.0
32.0
27.580
Normal Distribution
Mean
Standard Deviation
29.11
0.93
Edit Parameters
About 5% of the bootstrap means will be less than 27.58 minutes.

And in R ...

```
xqnorm(0.05, mean = 29.11, sd = 0.93)

##  P(X <= 27.5802861269351) = 0.05
##  P(X >  27.5802861269351) = 0.95
```




```
## [1] 27.58029
```

P-values Using a Normal

The mean commute time in the sample of 500 Atlanta commuters is 29.11 minutes. Is there evidence that the mean commute time for *all* Atlanta commuters is less than 30 minutes?

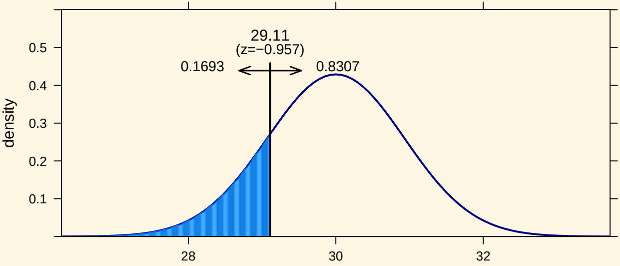
$$H_0 : \mu = 30$$
$$H_1 : \mu \neq 30$$

Suppose we can model the randomization distribution using a Normal with a standard error of 0.93. What should the mean be? Find the P-value.

In R ...

```
xpnorm(29.11, mean = 30, sd = 0.93)

##
## If X ~ N(30, 0.93), then
##
##  P(X <= 29.11) = P(Z <= -0.9569892) = 0.1692863
##  P(X >  29.11) = P(Z >  -0.9569892) = 0.8307137
```




```
## [1] 0.1692863
```

Outline

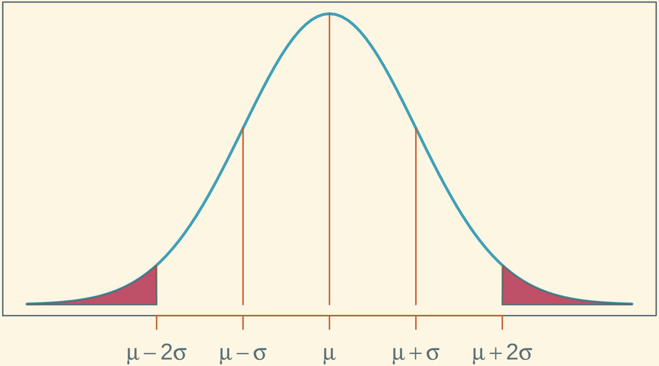
Standard Normal

CIs from a Standard Normal

P-values Using a Standard Normal

Quantiles of Normal Curves

The shape of a Normal is the same for all μ and σ. The mean is always at the peak; the "inflection points" are always $\mu + \sigma$ and $\mu - \sigma$, and 95% of the area is always between $\mu - 2\sigma$ and $\mu + 2\sigma$.



So, for proportions and quantiles, only "standard distances from the mean" (z-scores) matter!

What is a z-score?

The z-score for a point tells you how many standard deviations above the mean it is (negative = below)

$$Z = \frac{X - \mu}{\sigma} \qquad X = \sigma Z + \mu$$

If we relabel the x-axis of our density curve with a z-axis, we get what's called a **Standard Normal** distribution.

Normal and Standard Normal

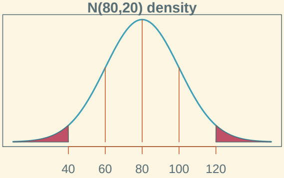


Figure: Left: Normal density with mean 80 and standard deviation 20. Right: Standard Normal (mean 0, standard deviation 1).

Example: Gestation Time

Dear Abby: You wrote that a woman is pregnant for 266 days. Who said so? I carried my baby for ten months and five days, and there is no doubt about it because I know the exact date my baby was conceived. My husband is in the Navy and it couldn't have possibly been conceived any other time because I saw him only once for an hour, and I didn't see him again until the day before the baby was born.

I don't drink or run around, and there is no way the baby isn't his, so please print a retraction about the 266-day carrying time because otherwise I'm in a lot of trouble.

San Diego Reader

Dear San Diego Reader: Some babies come early, some come late; yours came late.

Abby

Example: Gestation Time

Human gestation times in days are distributed approximately $\mathcal{N}(266, 16)$. The reader was pregnant for 305 days.

- What is that as a z-score?
- Use the raw score to find the reader's percentile.
- Use the z-score to find the reader's percentile.

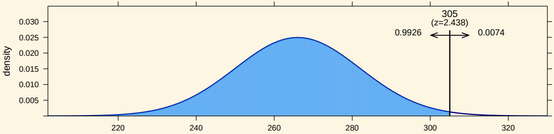
Solutions: Gestation Time

Human gestation times in days are distributed approximately $\mathcal{N}(266, 16)$. The reader was pregnant for 305 days.

$$z = \frac{X - \mu}{\sigma} = \frac{305 - 266}{16} = 2.4375$$

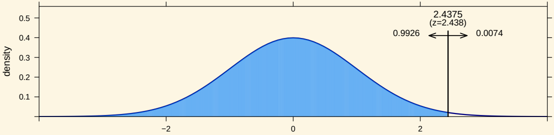
Solutions: Gestation Time

```
### Using the raw score, the percentile is given by xpnorm:
xpnorm(305, mean = 266, sd = 16, lower.tail = TRUE, verbose = FALSE)
```




```
## [1] 0.9926054
```

```
### When we use the z score, we locate it in the standard normal:
xpnorm(2.4375, mean = 0, sd = 1, lower.tail = TRUE, verbose = FALSE)
```



Outline

Standard Normal

CIs from a Standard Normal

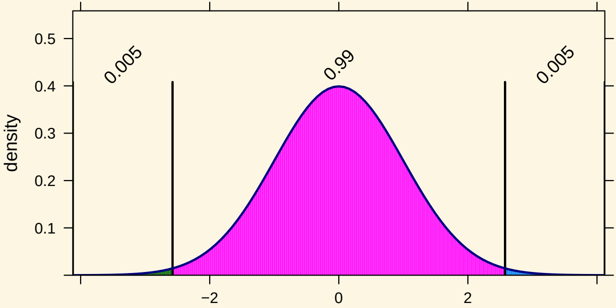
P-values Using a Standard Normal

Confidence Intervals from a Standard Normal

- We already know that Sample Statistic $\pm$ 2 SE yields an (approximately) 95% CI. What are the z-scores associated with these endpoints in the context of the bootstrap distribution?
- When the bootstrap distribution is Normal, the z-scores for a given confidence level are always the same.
  - 95%: $z \approx \pm 2$
  - 99%: ?
  - 90%: ?
- How can we find these using a standard normal?

Confidence Intervals from a Standard Normal

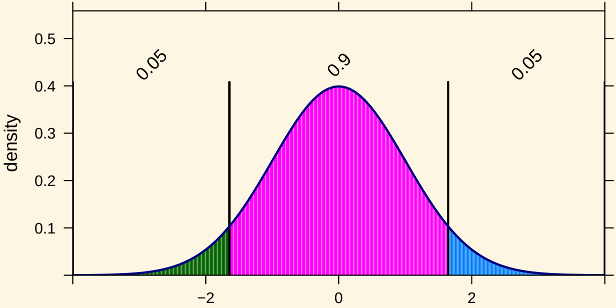
```
### Find the 0.005 and 0.995 quantiles of the standard Normal.
### These are the z-scores of the 99% confidence interval
### (within the bootstrap distribution)
xqnorm(c(0.005, 0.995), mean = 0, sd = 1, verbose = FALSE)
```




```
## [1] -2.575829  2.575829
```

Confidence Intervals from a Standard Normal

```
### Find the 0.05 and 0.95 quantiles of the standard Normal.
### These are the z-scores of the 90% confidence interval
### (within the bootstrap distribution)
xqnorm(c(0.05, 0.95), mean = 0, sd = 1, lower.tail = TRUE, verbose = FALSE)
```




```
## [1] -1.644854  1.644854
```

Z-score conversion

The relationship between the original scale and standardized scale is

$$Z = \frac{\text{Original} - \text{Distribution Mean}}{\text{Standard Deviation}}$$

Converting back to the original scale

If we find the z-scores of the CI endpoints, we can convert them to a confidence interval on the original scale.

$$\text{Endpoint(Original)} = \text{Distribution Mean} + Z \cdot \text{Standard Deviation}$$

Demo

Converting back to the original scale

If we find the z-scores of the CI endpoints, we can convert them to a confidence interval on the original scale.

$$\text{Endpoint(Original)} = \text{Distribution Mean} + Z \cdot \text{Standard Deviation}$$

CI Summary

To compute a confidence interval when the bootstrap distribution can be replaced by a Normal, use

$$\text{Endpoint} = \text{observed statistic} \pm Z^* \cdot \text{Bootstrap SE}$$

where Z^* is the Z-score of the endpoint appropriate for the confidence level, computed from a standard normal ($\mathcal{N}(0, 1)$).

Outline

Standard Normal

CIs from a Standard Normal

P-values Using a Standard Normal

P-values Using a Standard Normal

P-values from a Standard Normal

Computing P-values when the randomization distribution is Normal is the reverse process:

1. Convert the observed statistic to a z-score within the randomization distribution (i.e., using its mean and standard deviation).

$$Z_{observed} = \frac{\text{observed statistic} - \text{null parameter}}{\text{randomization SD}}$$

2. Find the relevant area beyond $Z_{observed}$ using a Standard Normal

Example: Sleep and Caffeine

Is mean number of words recalled different after sleep vs. caffeine?

$$H_0 : \mu_{\text{sleep}} - \mu_{\text{caffeine}} = 0$$
$$H_1 : \mu_{\text{sleep}} - \mu_{\text{caffeine}} \neq 0$$

Sample statistic: $\bar{x}_{\text{sleep}} - \bar{x}_{\text{caffeine}}$
P-value: Demo