z and t tests for the mean of a normal distribution
Confidence intervals for the mean
Binomial tests

Chapters 3.5.1–3.5.2, 3.3.2

Prof. Tesler

Math 283
Fall 2018

Sample mean: estimating μ from data

- A random variable has a normal distribution with mean $\mu = 500$ and standard deviation $\sigma = 100$, but those parameters are secret.
- We will study how to estimate their values as points or intervals and how to perform hypothesis tests on their values.

Parametric tests involving normal distribution

- z-test: σ known, μ unknown; testing value of μ
- t-test: σ, μ unknown; testing value of μ
- χ^2 test: σ unknown; testing value of σ
- Plus generalizations for comparing two or more random variables from different normal distributions:
  - Two-sample z and t tests: Comparing μ for two different normal variables.
  - F test: Comparing σ for two different normal variables.
  - ANOVA: Comparing μ between multiple normal variables.

Estimating parameters from data

Repeated measurements of X, which has mean μ and standard deviation σ

Basic experiment

1. Make independent measurements $x_1, \ldots, x_n$.
2. Compute the sample mean:
   $$m = \bar{x} = \frac{x_1 + \cdots + x_n}{n}$$
   The sample mean is a point estimate of μ; it just gives one number, without an indication of how far away it might be from μ.
3. Repeat the above with many independent samples, getting different sample means each time.

The long-term average of the sample means will be approximately

$$E(\overline{X}) = E\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{\mu+\cdots+\mu}{n} = \frac{n\mu}{n} = \mu$$

These estimates will be distributed with variance $\mathrm{Var}(\overline{X}) = \sigma^2/n$.

Sample variance s^2: estimating σ^2 from data

- **Data:** $1, 2, 12$
- **Sample mean:** $\bar{x} = \frac{1+2+12}{3} = 5$
- **Deviations of data from the sample mean, $x_i - \bar{x}$:** $1-5,\ 2-5,\ 12-5\ =\ -4,\ -3,\ 7$

- In this example, the deviations sum to $-4 - 3 + 7 = 0$.
- In general, the deviations sum to
  $$\left(\sum_{i=1}^{n} x_i\right) - n\bar{x} = 0$$
  since $\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n$.
- So, given any $n - 1$ of the deviations, the remaining one is determined. In this example, if you're told there are three deviations and given two of them,
  $$-4, \_\_, 7$$
  then the missing one has to be -3, so that they add up to 0.
- We say there are $n - 1$ degrees of freedom ($df = n - 1$).

Sample variance s^2: estimating σ^2 from data

- **Data:** $1, 2, 12$
- **Sample mean:** $\bar{x} = \frac{1+2+12}{3} = 5$
- **Deviations of data from the sample mean, $x_i - \bar{x}$:** $1-5,\ 2-5,\ 12-5\ =\ -4, -3, 7$

- Here, $df = 2$ and the sum of squared deviations is
$$ss = (-4)^2 + (-3)^2 + 7^2 = 16 + 9 + 49 = 74$$
- If the random variable X has true mean $\mu = 6$, the sum of squared deviations from $\mu = 6$ would be
$$(1-6)^2 + (2-6)^2 + (12-6)^2 = (-5)^2 + (-4)^2 + 6^2 = 77$$
- $\sum_{i=1}^{n} (x_i - y)^2$ is minimized at $y = \bar{x}$, so ss underestimates $\sum_{i=1}^{n} (x_i - \mu)^2$.

Sample variance: estimating σ^2 from data

Definitions

Sum of squared deviations: $ss = \sum_{i=1}^{n} (x_i - \bar{x})^2$

Sample variance: $s^2 = \frac{ss}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Sample standard deviation: $s = \sqrt{s^2}$

- s^2 turns out to be an unbiased estimate of σ^2: $E(S^2) = \sigma^2$.
- For the sake of demonstration, let $u^2 = \frac{ss}{n} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$.
- Although u^2 is the MLE of σ^2 for the normal distribution, it is biased: $E(U^2) = \frac{n-1}{n} \sigma^2$.
- This is because $\sum_{i=1}^{n} (x_i - \bar{x})^2$ underestimates $\sum_{i=1}^{n} (x_i - \mu)^2$.

Estimating μ and σ^2 from sample data (secret: $\mu = 500$, $\sigma = 100$)

Exp. #	x_1	x_2	x_3	x_4	x_5	x_6	$\bar{x}$	$s^2 = ss/5$	$u^2 = ss/6$
1	550	600	450	400	610	500	518.33	7016.67	5847.22
2	500	520	370	520	480	440	471.67	3376.67	2813.89
3	470	530	610	370	350	710	506.67	19426.67	16188.89
4	630	620	430	470	500	470	520.00	7120.00	5933.33
5	690	470	500	410	510	360	490.00	12840.00	10700.00
6	450	490	500	380	530	680	505.00	10030.00	8358.33
7	510	370	480	400	550	530	473.33	5306.67	4422.22
8	420	330	540	460	630	390	461.67	11736.67	9780.56
9	570	430	470	520	450	560	500.00	3440.00	2866.67
10	260	530	330	490	530	630	461.67	19296.67	16080.56
						Average	490.83	9959.00	8299.17

- We used $n = 6$, repeated for 10 trials, to fit the slide, but larger values would be better in practice.
- **Average of $\bar{x}$:** $490.83 \approx \mu = 500$ ✓
- **Average of $s^2 = ss/5$:** $9959.00 \approx \sigma^2 = 10000$ ✓
- **Average of $u^2 = ss/6$:** $8299.17 \approx \frac{n-1}{n}\sigma^2 = 8333.33$ ✗

Proof that denominator $n-1$ makes s^2 unbiased

- Expand the $i=1$ term of $SS = \sum_{i=1}^{n}(X_i - \overline{X})^2$:

$$E((X_1-\overline{X})^2) = E({X_1}^2) + E(\overline{X}^2) - 2E(X_1\overline{X})$$

- $\mathrm{Var}(X) = E(X^2) - E(X)^2 \quad \Rightarrow \quad E(X^2) = \mathrm{Var}(X) + E(X)^2.$ So

$$E({X_1}^2) = \sigma^2 + \mu^2 \qquad E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$$

- Cross-term:

$$\begin{aligned} E(X_1\overline{X}) &= \frac{E({X_1}^2) + E(X_1)E(X_2) + \cdots + E(X_1)E(X_n)}{n} \\ &= \frac{(\sigma^2+\mu^2) + (n-1)\mu^2}{n} = \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

- Total for $i=1$ term:

$$E((X_1-\overline{X})^2) = (\sigma^2+\mu^2) + \left(\frac{\sigma^2}{n}+\mu^2\right) - 2\left(\frac{\sigma^2}{n}+\mu^2\right) = \frac{n-1}{n}\sigma^2$$

Proof that denominator $n-1$ makes s^2 unbiased

- Similarly, every term of $SS = \sum_{i=1}^{n} (X_i - \overline{X})^2$ has

$$E\left((X_i - \overline{X})^2\right) = \frac{n-1}{n}\sigma^2$$

- The total is

$$E(SS) = (n-1)\sigma^2$$

- Thus we must divide SS by $n-1$ instead of n to get an unbiased estimator of σ^2.

Data

Exp. #	Values $x_1, \ldots, x_6$	Sample mean $\bar{x}$	Sample Var. s^2	Sample SD s
#1	650, 510, 470, 570, 410, 370	496.67	10666.67	103.28
#2	510, 420, 520, 360, 470, 530	468.33	4456.67	66.76
#3	470, 380, 480, 320, 430, 490	428.33	4456.67	66.76

Suppose we do the "sample 6 scores" experiment a few times and get these values. We'll test

$$H_0: \quad \mu = 500 \qquad \text{vs.} \qquad H_1: \quad \mu \neq 500$$

for each of these under the assumption that the data comes from a normal distribution, with significance level $\alpha = 5\%$.

Number of standard deviations $\bar{x}$ is away from μ when $\mu = 500$ and $\sigma = 100$, for sample mean of $n = 6$ points

Number of standard deviations if σ is known:

The z-score of $\bar{x}$ is

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{x} - 500}{100 / \sqrt{6}}$$

Estimating number of standard deviations if σ is unknown:

The t-*score* of $\bar{x}$ is

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{\bar{x} - 500}{s / \sqrt{6}}$$

- It uses sample standard deviation s in place of σ.
- Note that s is computed from the same data as $\bar{x}$.
  The data feeds into the numerator and denominator of t.
- t has the same degrees of freedom as s; here, $df = n - 1 = 5$.
- As random variable: T_5 (T distribution with 5 degrees of freedom).

Number of standard deviations $\bar{x}$ is away from μ

Data

Exp. #	Values $x_1, \ldots, x_6$	Sample mean $\bar{x}$	Sample Var. s^2	Sample SD s
#1	650, 510, 470, 570, 410, 370	496.67	10666.67	103.28
#2	510, 420, 520, 360, 470, 530	468.33	4456.67	66.76
#3	470, 380, 480, 320, 430, 490	428.33	4456.67	66.76

#1: $z = \dfrac{496.67 - 500}{100/\sqrt{6}} \approx -.082$ $\quad t = \dfrac{496.67 - 500}{103.28/\sqrt{6}} \approx -.079$ **Close**

#2: $z = \dfrac{468.33 - 500}{100/\sqrt{6}} \approx -.776$ $\quad t = \dfrac{468.33 - 500}{66.76/\sqrt{6}} \approx -1.162$ **Far**

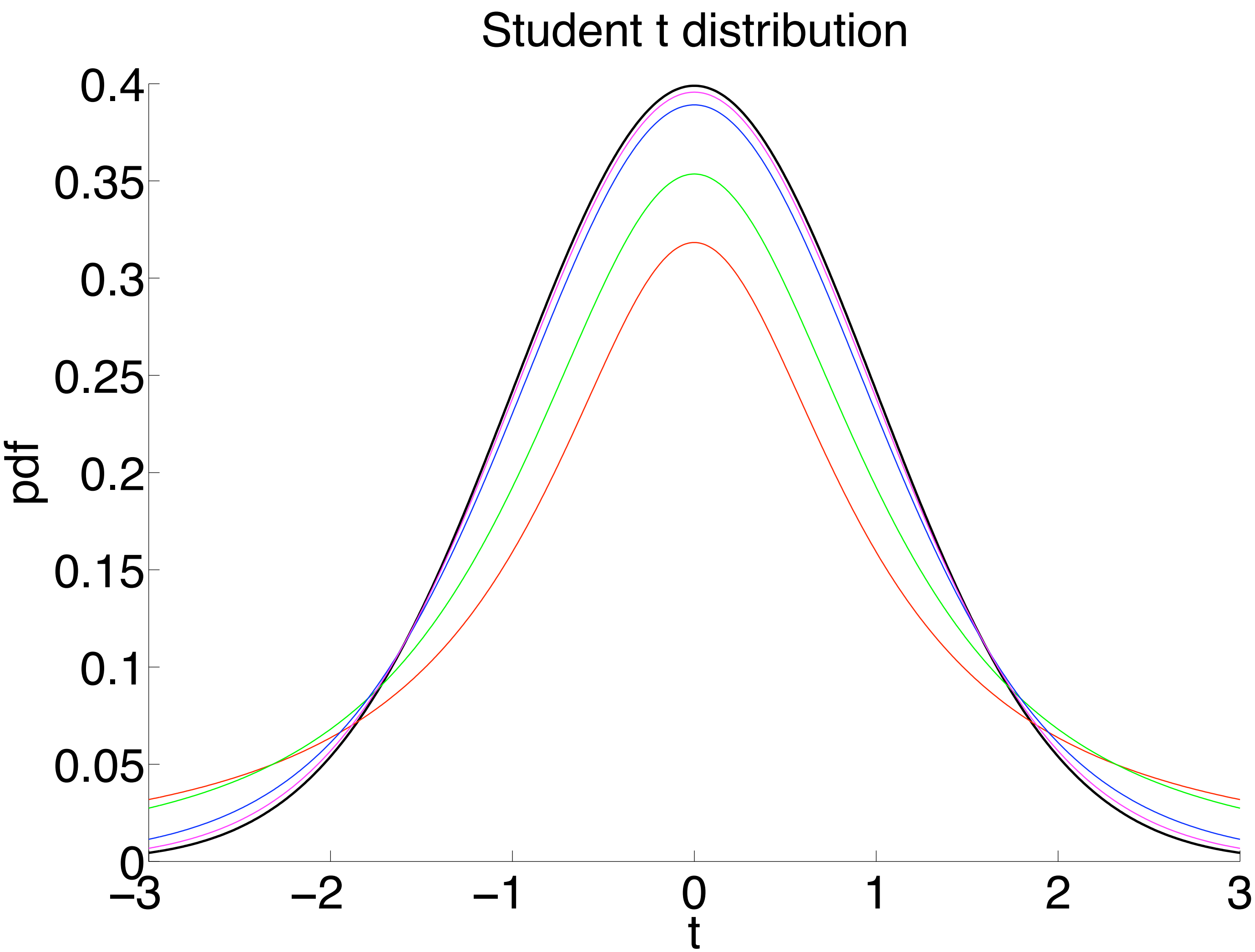
#3: $z = \dfrac{428.33 - 500}{100/\sqrt{6}} \approx -1.756$ $\quad t = \dfrac{428.33 - 500}{66.76/\sqrt{6}} \approx -2.630$ **Far**

Student t distribution

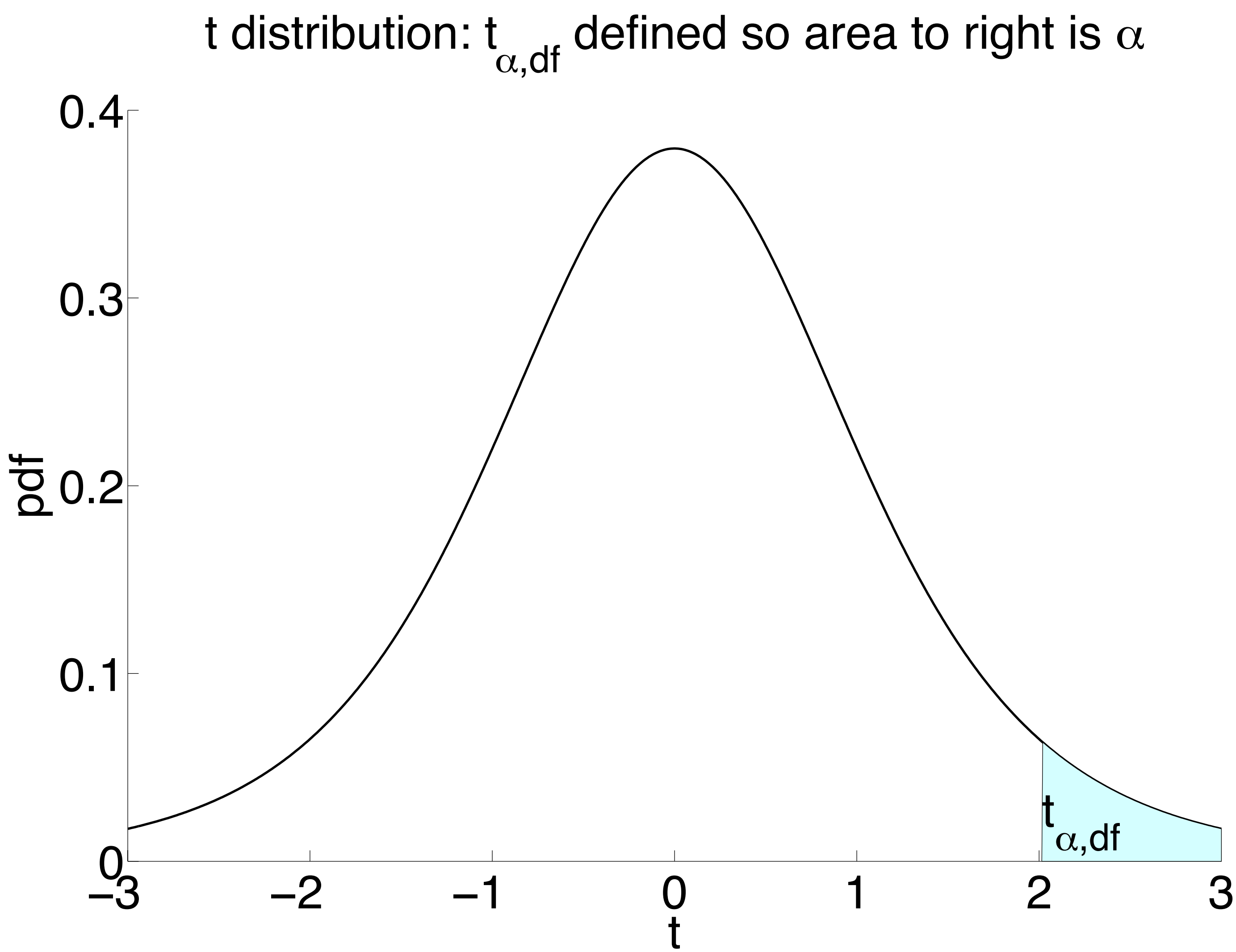
- In $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, the numerator depends on $x_1, \ldots, x_n$ while the denominator is constant.
  In $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$, both the numerator and denominator depend on x_i's.
- Random variable T_{n-1} has the t-distribution with $n-1$ degrees of freedom ($d.f. = n-1$).
- The pdf is still symmetric and "bell-shaped," *but not the same "bell" as the normal distribution.*
- Degrees of freedom $d.f.=n-1$ match here and in the s^2 formula.
- As degrees of freedom rises, the pdf gets closer to the standard normal pdf. They are really close for $d.f. \geqslant 30$.
- Developed by William Gosset (1908) while doing statistical tests on yeast at Guinness Brewery in Ireland. He found the z-test was inaccurate for small n. He published under pseudonym "Student."

Student t distribution

The curves from bottom to top (at $t = 0$) are for $d.f. = 1, 2, 10, 30$, and the top one is the standard normal curve:



Critical values of z or t

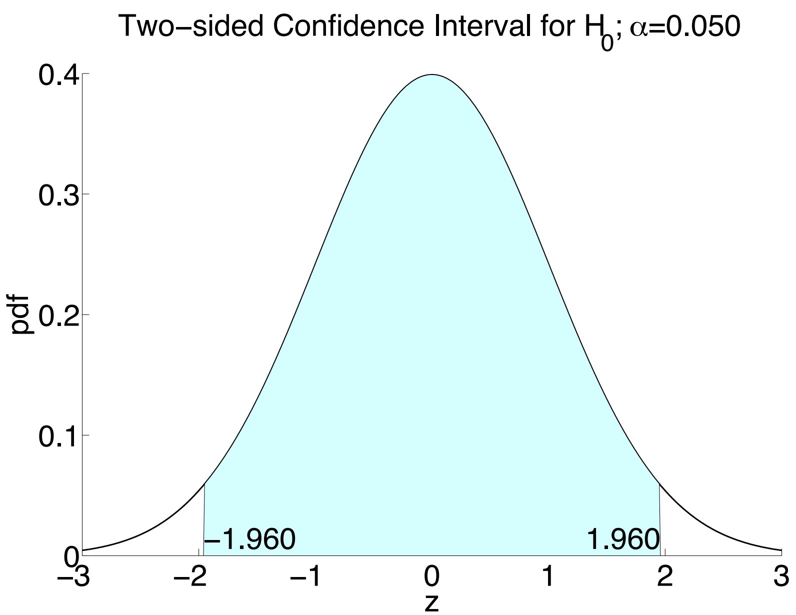


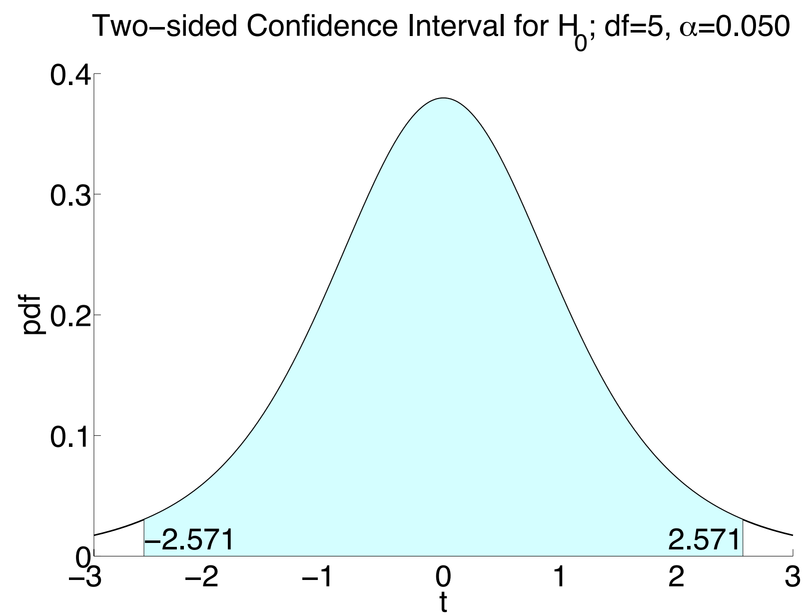
The values of z and t that put area α at the right are z_α and $t_{\alpha,df}$:

$$P(Z \geqslant z_\alpha) = \alpha \qquad\qquad P(T_{df} \geqslant t_{\alpha,df}) = \alpha$$

Computing critical values of z or t with Matlab

- We'll use significance level $\alpha = 5\%$ and $n = 6$ data points, so $df = n - 1 = 5$ for t.
- We want areas $\alpha/2 = 0.025$ on the left and right and $1 - \alpha = 0.95$ in the center.
- The Matlab and R functions shown below use areas to the left.
  Therefore, to get area $.025$ on the right, look up the cutoff for area $.975$ on the left.





	Matlab		R	
$-z_{0.025}$ =	`norminv(.025)`	=	`qnorm(.025)`	$= -1.96$
$z_{0.025}$ =	`norminv(.975)`	=	`qnorm(.975)`	$= 1.96$
	`normcdf(-1.96)`	=	`pnorm(-1.96)`	$= 0.025$
	`normcdf(1.96)`	=	`pnorm(1.96)`	$= 0.975$
	`normpdf(-1.96)`	=	`dnorm(-1.96)`	$= 0.0584$
	`normpdf(1.96)`	=	`dnorm(1.96)`	$= 0.0584$

	Matlab		R	
$-t_{0.025,5}$ =	`tinv(.025,5)`	=	`qt(.025,5)`	$= -2.5706$
$t_{0.025,5}$ =	`tinv(.975,5)`	=	`qt(.975,5)`	$= 2.5706$
	`tcdf(-2.5706,5)`	=	`pt(-2.5706,5)`	$= 0.0250$
	`tcdf(2.5706,5)`	=	`pt(2.5706,5)`	$= 0.9750$
	`tpdf(-2.5706,5)`	=	`dt(-2.5706,5)`	$= 0.0303$
	`tpdf(2.5706,5)`	=	`dt(2.5706,5)`	$= 0.0303$

Hypothesis tests for μ

Test H_0: $\mu = 500$ vs. H_1: $\mu \neq 500$ at significance level $\alpha = .05$

Exp. #	Data $x_1, \ldots, x_6$	$\bar{x}$	s^2	s
#1	650, 510, 470, 570, 410, 370	496.67	10666.67	103.28
#2	510, 420, 520, 360, 470, 530	468.33	4456.67	66.76
#3	470, 380, 480, 320, 430, 490	428.33	4456.67	66.76

When σ is known (say $\sigma = 100$)

Reject H_0 when $|z| \geqslant z_{\alpha/2} = z_{.025} = 1.96$.

#1: $z = -.082$, $|z| < 1.96$ so accept H_0.

#2: $z = -.776$, $|z| < 1.96$ so accept H_0.

#3: $z = -1.756$, $|z| < 1.96$ so accept H_0.

When σ is not known, but is estimated by s

Reject H_0 when $|t| \geqslant t_{\alpha/2,n-1} = t_{.025,5} = 2.5706$.

#1: $t = -.079$, $|t| < 2.5706$ so accept H_0.

#2: $t = -1.162$, $|t| < 2.5706$ so accept H_0.

#3: $t = -2.630$, $|t| \geqslant 2.5706$ so reject H_0.

One-sided hypothesis test: Left-sided critical region

$H_0 : \mu = 500$ vs. $H_1 : \mu < 500$ at significance level $\alpha = 5\%$

The cutoffs to put 5% of the area at the left are

	Matlab		R	
$-z_{0.05} =$	`norminv(0.05)`	$=$	`qnorm(0.05)`	$= -1.6449$
$-t_{0.05,5} =$	`tinv(0.05,5)`	$=$	`qt(0.05,5)`	$= -2.0150$

When σ is known (say $\sigma = 100$)

Reject H_0 when $z \leqslant -z_\alpha = -z_{.05} = -1.6449$:

#1: $z = -.082$, $z > -1.6449$ so accept H_0.

#2: $z = -.776$, $z > -1.6449$ so accept H_0.

#3: $z = -1.756$, $z \leqslant -1.6449$ so reject H_0.

When σ is not known, but is estimated by s

Reject H_0 when $t \leqslant -t_{\alpha,n-1} = -t_{.05,5} = -2.0150$.

#1: $t = -.079$, $t > -2.0150$ so accept H_0.

#2: $t = -1.162$, $t > -2.0150$ so accept H_0.

#3: $t = -2.630$, $t \leqslant -2.0150$ so reject H_0.

One-sided hypothesis test: Right-sided critical region

$H_0 : \mu = 500$ vs. $H_1 : \mu > 500$ at significance level $\alpha = 5\%$
The cutoffs to put 5% of the area at the right are

	Matlab		**R**	
$z_{0.05} =$	`norminv(0.95)`	$=$	`qnorm(0.95)`	$= 1.6449$
$t_{0.05,5} =$	`tinv(0.95,5)`	$=$	`qt(0.95,5)`	$= 2.0150$

When σ is known (say $\sigma = 100$)

Reject H_0 when $z \geqslant z_\alpha = z_{.05} = 1.6449$:

#1: $z = -.082$, $z < 1.6449$ so accept H_0.
#2: $z = -.776$, $z < 1.6449$ so accept H_0.
#3: $z = -1.756$, $z < 1.6449$ so accept H_0.

When σ is not known, but is estimated by s

Reject H_0 when $t \geqslant t_{\alpha,n-1} = t_{.05,5} = 2.0150$.

#1: $t = -.079$, $t < 2.0150$ so accept H_0.
#2: $t = -1.162$, $t < 2.0150$ so accept H_0.
#3: $t = -2.630$, $t < 2.0150$ so accept H_0.

Z-tests using P-values, data set #2 ($z = -0.776$)

(a) $H_0: \mu = 500$
$H_1: \mu > 500$

$$\begin{aligned} P &= P(Z \geqslant -0.776) \\ &= 1 - \Phi(-0.776) \\ &= 1 - .2189 \\ &= .7811 \end{aligned}$$

R: $1-$pnorm$(-.776)$

Matlab: $1-$normcdf$(-.776)$

(b) $H_0: \mu = 500$
$H_1: \mu < 500$

$$\begin{aligned} P &= P(Z \leqslant -0.776) \\ &= \Phi(-0.776) \\ &= .2189 \end{aligned}$$

R: pnorm$(-.776)$

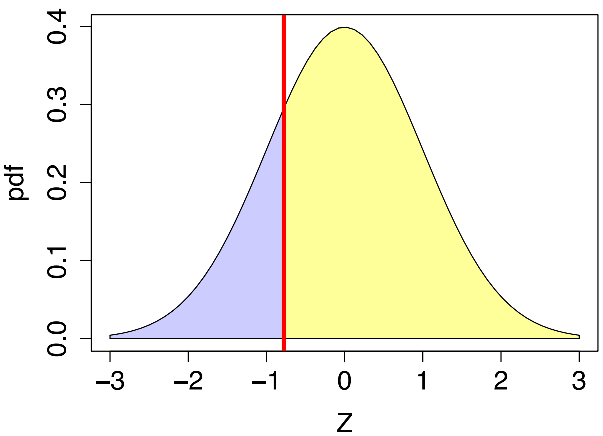
Matlab: normcdf$(-.776)$

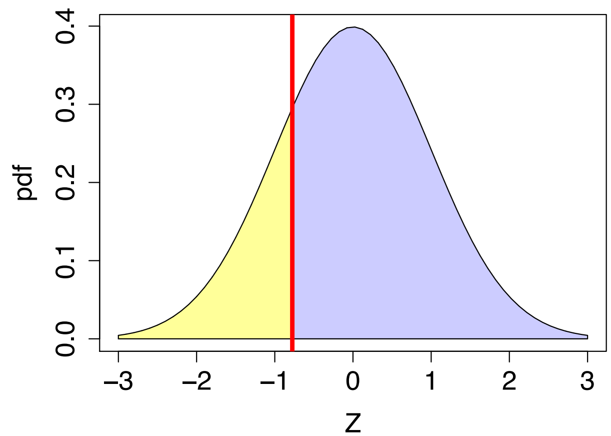
(c) $H_0: \mu = 500$
$H_1: \mu \neq 500$

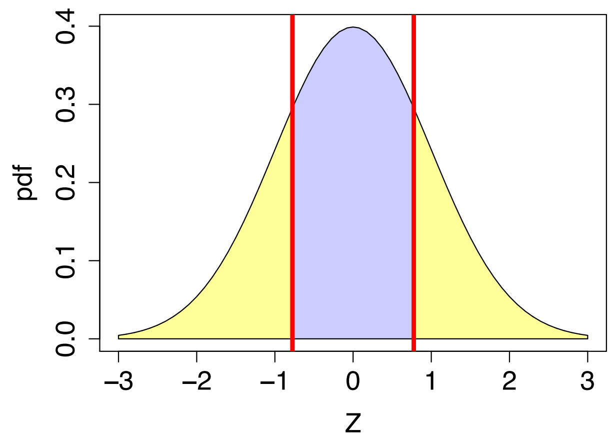
$$\begin{aligned} P &= P(|Z| \geqslant 0.776) \\ &= 2P(Z \geqslant 0.776) \\ &= 2(.2189) \\ &= .4377 \end{aligned}$$

R: $2*$pnorm$(-.776)$

Matlab: $2*$normcdf$(-.776)$







Supports H_0 better
Supports H_1 better
Observed z=−0.776

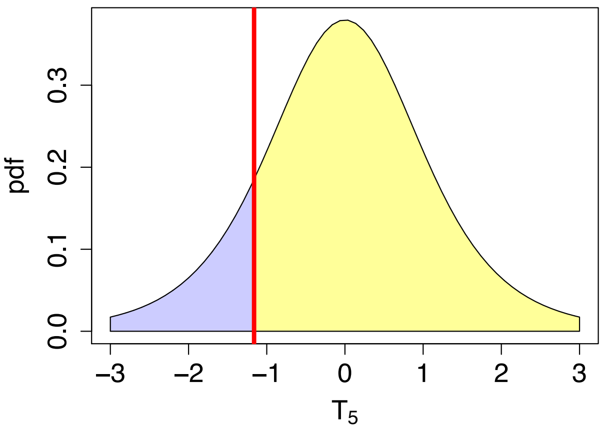
In each case, $P > \alpha = 0.05$, so accept H_0.

T-tests using P-values, data set #2 ($t = -1.162$, $df = 5$)

(a) H_0: $\mu = 500$
H_1: $\mu > 500$

$$P = P(T_5 \geqslant -1.162) = 1 - P(T_5 < -1.162) = 1 - .1488 = .8512$$

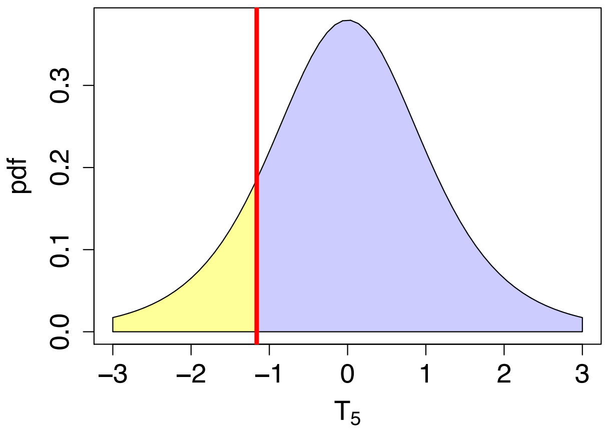
R: 1−pt(−1.162,5)
Matlab: 1−tcdf(−1.162,5)



(b) H_0: $\mu = 500$
H_1: $\mu < 500$

$$P = P(T_5 \leqslant -1.162) = .1488$$

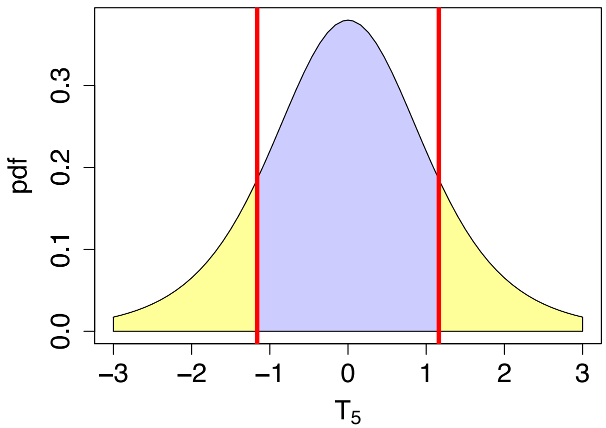
pt(−1.162,5)
tcdf(−1.162,5)



(c) H_0: $\mu = 500$
H_1: $\mu \neq 500$

$$P = P(|T_5| \geqslant 1.162) = 2P(T_5 \leqslant -1.162) = 2(.1488) = .2977$$

2∗pt(−1.162,5)
2∗tcdf(−1.162,5)



- Supports H_0 better
- Supports H_1 better
- Observed t=−1.162

In each case, $P > \alpha = 0.05$, so accept H_0.

(2-sided) confidence intervals for estimating μ from $\bar{x}$

(Chapter 3.3.2)

- If our data comes from a normal distribution with known σ then 95% of the time, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ should lie between ± 1.96.
- Solve for bounds on μ from the upper limit on Z:

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant 1.96 \quad \Leftrightarrow \quad \bar{x}-\mu \leqslant 1.96\frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad \bar{x}-1.96\frac{\sigma}{\sqrt{n}} \leqslant \mu$$

  Notice the 1.96 turned into -1.96 and we get a lower limit on μ.

- Also solve for an upper bound on μ from the lower limit on Z:

$$-1.96 \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \Leftrightarrow \quad -1.96\frac{\sigma}{\sqrt{n}} \leqslant \bar{x}-\mu \quad \Leftrightarrow \quad \mu \leqslant \bar{x}+1.96\frac{\sigma}{\sqrt{n}}$$

- Together, $\quad \bar{x}-1.96\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x}+1.96\frac{\sigma}{\sqrt{n}}$
- In the long run, μ is contained in approximately 95% of intervals

$$\left(\bar{x}-1.96\frac{\sigma}{\sqrt{n}},\ \bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

  This interval is called a *confidence interval*.

2-sided $(100-\alpha)\%$ confidence interval for the mean

When σ is known

$$\left(\bar{x} - \frac{z_{\alpha/2}}{\sqrt{n}}\sigma \quad , \quad \bar{x} + \frac{z_{\alpha/2}}{\sqrt{n}}\sigma\right)$$

95% confidence interval ($\alpha = 5\% = 0.05$) with $\sigma = 100$, $z_{.025} = 1.96$:

$$\left(\bar{x} - \frac{1.96(100)}{\sqrt{n}} \quad , \quad \bar{x} + \frac{1.96(100)}{\sqrt{n}}\right)$$

Other commonly used percentages:

99% CI: use ± 2.58 instead of ± 1.96. 90% CI: use ± 1.64.

For demo purposes: 75% CI: use ± 1.15.

When σ is not known, but is estimated by s

$$\left(\bar{x} - \frac{t_{\alpha/2,n-1}}{\sqrt{n}}s \quad , \quad \bar{x} + \frac{t_{\alpha/2,n-1}}{\sqrt{n}}s\right)$$

A 95% confidence interval when $n = 6$ is $\left(\bar{x} - \frac{2.5706s}{\sqrt{n}} \quad , \quad \bar{x} + \frac{2.5706s}{\sqrt{n}}\right)$.

The cutoff 2.5706 depends on α **and n**, so would change if n changes.

95% confidence intervals for μ

Exp. #	Data $x_1, \ldots, x_6$	$\bar{x}$	s^2	s
#1	650, 510, 470, 570, 410, 370	496.67	10666.67	103.28
#2	510, 420, 520, 360, 470, 530	468.33	4456.67	66.76
#3	470, 380, 480, 320, 430, 490	428.33	4456.67	66.76

When σ known (say $\sigma = 100$), use normal distribution

#1: $\left(496.67 - \frac{1.96(100)}{\sqrt{6}}, 496.67 + \frac{1.96(100)}{\sqrt{6}}\right) = (416.65, 576.69)$

#2: $\left(468.33 - \frac{1.96(100)}{\sqrt{6}}, 468.33 + \frac{1.96(100)}{\sqrt{6}}\right) = (388.31, 548.35)$

#3: $\left(428.33 - \frac{1.96(100)}{\sqrt{6}}, 428.33 + \frac{1.96(100)}{\sqrt{6}}\right) = (348.31, 508.35)$

When σ not known, estimate σ by s and use t-distribution

#1: $\left(496.67 - \frac{2.5706(103.28)}{\sqrt{6}}, 496.67 + \frac{2.5706(103.28)}{\sqrt{6}}\right) = (388.28, 605.06)$

#2: $\left(468.33 - \frac{2.5706(66.76)}{\sqrt{6}}, 468.33 + \frac{2.5706(66.76)}{\sqrt{6}}\right) = (398.27, 538.39)$

#3: $\left(428.33 - \frac{2.5706(66.76)}{\sqrt{6}}, 428.33 + \frac{2.5706(66.76)}{\sqrt{6}}\right) = (358.27, 498.39)$
(missing 500)

Confidence intervals

$\sigma = 100$ known, $\mu = 500$ unknown, $n = 6$ points per trial, 20 trials

Confidence intervals w/o $\mu = 500$ are marked *(393.05,486.95)*.

Trial #	x_1	x_2	x_3	x_4	x_5	x_6	$m = \bar{x}$	75% conf. int.	95% conf. int.
1	720	490	660	520	390	390	528.33	(481.38,575.28)	(448.32,608.35)
2	380	260	390	630	540	440	440.00	*(393.05,486.95)*	(359.98,520.02)
3	800	450	580	520	650	390	565.00	*(518.05,611.95)*	(484.98,645.02)
4	510	370	530	290	460	540	450.00	*(403.05,496.95)*	(369.98,530.02)
5	580	500	540	540	340	340	473.33	(426.38,520.28)	(393.32,553.35)
6	500	490	480	550	390	450	476.67	(429.72,523.62)	(396.65,556.68)
7	530	680	540	510	520	590	561.67	*(514.72,608.62)*	(481.65,641.68)
8	480	600	520	600	520	390	518.33	(471.38,565.28)	(438.32,598.35)
9	340	520	500	650	400	530	490.00	(443.05,536.95)	(409.98,570.02)
10	460	450	500	360	600	440	468.33	(421.38,515.28)	(388.32,548.35)
11	540	520	360	500	520	640	513.33	(466.38,560.28)	(433.32,593.35)
12	440	420	610	530	490	570	510.00	(463.05,556.95)	(429.98,590.02)
13	520	570	430	320	650	540	505.00	(458.05,551.95)	(424.98,585.02)
14	560	380	440	610	680	460	521.67	(474.72,568.62)	(441.65,601.68)
15	460	590	350	470	420	740	505.00	(458.05,551.95)	(424.98,585.02)
16	430	490	370	350	360	470	411.67	*(364.72,458.62)*	*(331.65,491.68)*
17	570	610	460	410	550	510	518.33	(471.38,565.28)	(438.32,598.35)
18	380	540	570	400	360	500	458.33	(411.38,505.28)	(378.32,538.35)
19	410	730	480	600	270	320	468.33	(421.38,515.28)	(388.32,548.35)
20	490	390	450	610	320	440	450.00	*(403.05,496.95)*	(369.98,530.02)

Confidence intervals

$\sigma = 100$ known, $\mu = 500$ unknown, $n = 6$ points per trial, 20 trials

- In the 75% confidence interval column, 14 out of 20 (70%) intervals contain the mean ($\mu = 500$).
This is close to 75%.

- In the 95% confidence interval column, 19 out of 20 (95%) intervals contain the mean ($\mu = 500$).
This is exactly 95% (though if you do it 20 more times, it wouldn't necessarily be exactly 19 the next time).

- A k% confidence interval means if we repeat the experiment a lot of times, *approximately* k% of the intervals will contain μ.
It is *not* a guarantee that exactly k% will contain it.

- *Note:* If you really don't know the true value of μ, you can't actually mark the intervals that do or don't contain it.

Confidence intervals — choosing n

Data:	380, 260, 390, 630, 540, 440
Sample mean:	$\bar{x} = \frac{380+260+390+630+540+440}{6} = 440$
σ:	We assume $\sigma = 100$ is known
95% CI half-width:	$1.96\frac{\sigma}{\sqrt{n}} = \frac{(1.96)(100)}{\sqrt{6}} \approx 80.02$
95% CI:	$(440 - 80.02, 440 + 80.02) = (359.98, 520.02)$

- To get a narrower 95% confidence interval, say mean ± 10, solve for n making the half-width $\leqslant$ 10:

$$1.96\frac{\sigma}{\sqrt{n}} \leqslant 10 \qquad n \geqslant \left(\frac{1.96\sigma}{10}\right)^2 = \left(\frac{1.96(100)}{10}\right)^2 = 384.16 \qquad n \geqslant 385$$

One-sided confidence intervals

- In a two-sided 95% confidence interval, we excluded the highest and lowest 2.5% of values and keep the middle 95%. One-sided removes the whole 5% from one side.

One-sided to the right: remove highest (right) 5% values of Z

$$P(Z \leqslant z_{.05}) = P(Z \leqslant 1.64) = .95$$

95% of experiments have $\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant 1.64$ so $\mu \geqslant \bar{x} - 1.64\dfrac{\sigma}{\sqrt{n}}$

So the one-sided (right) 95% CI for μ is $(\bar{x} - 1.64\frac{\sigma}{\sqrt{n}}, \infty)$

One-sided to the left: remove lowest (left) 5% of values of Z

$$P(-z_{.05} \leqslant Z) = P(-1.64 \leqslant Z) = .95$$

The one-sided (left) 95% CI for μ is $(-\infty, \bar{x} + 1.64\frac{\sigma}{\sqrt{n}})$

- If σ is estimated by s, use the t distribution cutoffs instead.

Hypothesis tests for the binomial distribution parameter p

Consider a coin with probability p of heads, $1 - p$ of tails.
Warning: do not confuse this with the P from P-values.

Two-sided hypothesis test: Is the coin fair?

Null hypothesis: H_0: $p = .5$ ("coin is fair")
Alternative hypothesis: H_1: $p \neq .5$ ("coin is not fair")

Draft of decision procedure

- Flip a coin 100 times.
- Let X be the number of heads.
- If X is "close" to 50 then it's fair, and otherwise it's not fair.

How do we quantify "close"?

Decision procedure — confidence interval

How do we quantify "close"?

Normal approximation to binomial $n = 100$, $p = 0.5$

$$\mu = np = 100(.5) = 50$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{100(.5)(1-.5)} = \sqrt{25} = 5$$

Check that it's OK to use the normal approximation:

$$\mu - 3\sigma = 50 - 15 = 35 > 0$$

$$\mu + 3\sigma = 50 + 15 = 65 < 100 \quad \text{so it is OK.}$$

$\approx$ 95% acceptance region

$$\begin{aligned}(\mu - 1.96\sigma, \mu + 1.96\sigma) &= (50 - 1.96 \cdot 5 \,,\, 50 + 1.96 \cdot 5) \\ &= (40.2 \,,\, 59.8)\end{aligned}$$

Decision procedure

Hypotheses

Null hypothesis: H_0: $p = .5$ ("coin is fair")
Alternative hypothesis: H_1: $p \neq .5$ ("coin is not fair")

Decision procedure

- Flip a coin 100 times.
- Let X be the number of heads.
- If $40.2 < X < 59.8$ then accept H_0; otherwise accept H_1.

Significance level: $\approx 5\%$

If H_0 is true (coin is fair), this procedure will give the wrong answer (H_1) about 5% of the time.

Measuring Type I error (a.k.a. *Significance Level*)

H_0 is the true state of nature, but we mistakenly reject H_0 / accept H_1

- If this were truly the normal distribution, the Type I error would be $\alpha = .05 = 5\%$ because we made a 95% confidence interval.
- However, the normal distribution is just an approximation; it's really the binomial distribution. So:

$$\begin{aligned}
\alpha &= P(\text{accept } H_1 | H_0 \text{ true}) \\
&= 1 - P(\text{accept } H_0 | H_0 \text{ true}) \\
&= 1 - P(40.2 < X < 59.8 \,|\, \text{binomial with } p = .5) \\
&= 1 - .9431120664 = 0.0568879336 \approx 5.7\%
\end{aligned}$$

$$\begin{aligned}
P(40.2 < X < 59.8 \,|\, p = .5) &= \sum_{k=41}^{59} \binom{100}{k} (.5)^k (1 - .5)^{100-k} \\
&= .9431120664
\end{aligned}$$

- So it's a 94.3% confidence interval and the Type I error rate is $\alpha = 5.7\%$.

Measuring Type II error

H_1 is the true state of nature but we mistakenly accept H_0 / reject H_1

- If $p = .7$, the test will probably detect it.
- If $p = .51$, the test will frequently conclude H_0 is true when it shouldn't, giving a high Type II error rate.
- If this were a game in which you won \$1 for each heads and lost \$1 for tails, there would be an incentive to make a biased coin with p just above $.5$ (such as $p = .51$) so it would be hard to detect.

Measuring Type II error

Exact Type II error for $p = .7$ using binomial distribution

- $$\begin{aligned}\beta &= P(\text{Type II error with } p = .7)\\ &= P(\text{Accept } H_0 \,|\, X \text{ is binomial, } p = .7)\\ &= P(40.2 < X < 59.8 \,|\, X \text{ is binomial, } p = .7)\\ &= \sum_{k=41}^{59} \binom{100}{k} (.7)^k (.3)^{100-k} = .0124984 \approx 1.25\%.\end{aligned}$$

- **When $p = 0.7$, the Type II error rate, β, is 1.25%:**
  $\approx 1.25\%$ of decisions made with a biased coin (specifically biased at $p = 0.7$) would incorrectly conclude H_0 (the coin is fair, $p = 0.5$).

- Since $H_1 : p \neq .5$ includes many different values of p, the Type II error rate depends on the specific value of p.

Measuring Type II error

Approximate Type II error using normal distribution

- $\mu = np = 100(.7) = 70$
- $\sigma = \sqrt{np(1-p)} = \sqrt{100(.7)(.3)} = \sqrt{21}$
- $$\begin{aligned}
\beta &= P(\text{Accept } H_0 \,|\, H_1 \text{ true: } X \text{ binomial with } n = 100,\ p = .7) \\
&\approx P(40.2 < X < 59.8 \,|\, X \text{ is normal with } \mu = 70,\ \sigma = \sqrt{21}) \\
&= P\left(\frac{40.2 - 70}{\sqrt{21}} < \frac{X - 70}{\sqrt{21}} < \frac{59.8 - 70}{\sqrt{21}}\right) \\
&\approx P(-6.5029 < Z < -2.2258) \qquad (\approx \text{ due to rounding}) \\
&= \Phi(-2.2258) - \Phi(-6.5029) \\
&\approx .0130 - .0000 = .0130 = 1.30\%
\end{aligned}$$

which is close to the exact value, $\approx 1.25\%$.

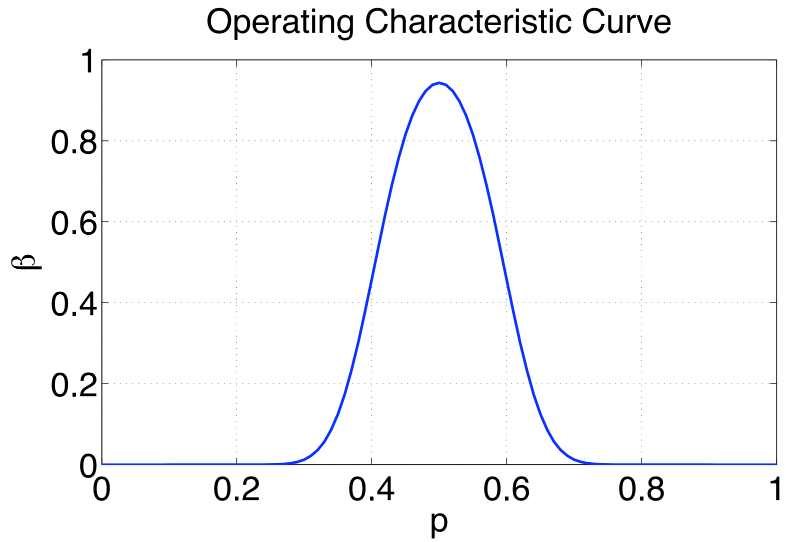
Power curve

- The decision procedure is "Flip a coin 100 times, let X be the number of heads, and accept H_0 if $40.2 < X < 59.8$".
- Plot the Type II error rate as a function of p:

$$\beta = \beta(p) = \sum_{k=41}^{59} \binom{100}{k} p^k (1-p)^{100-k}$$

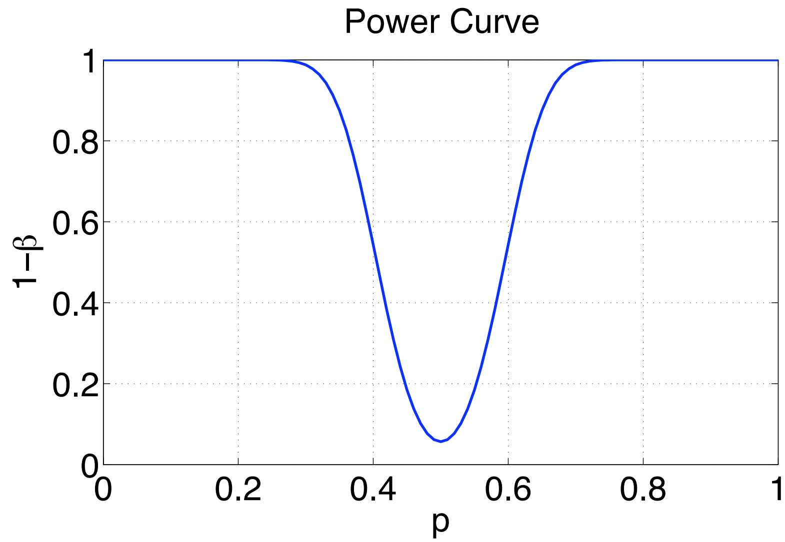
Type II Error:

$\beta = P(\textbf{Accept } H_0 \,|\, H_1 \textbf{ true})$



Correct detection of H_1:
Power = Sensitivity =

$1 - \beta = P(\textbf{Accept } H_1 \,|\, H_1 \textbf{ true})$



Choosing n to control Type I and II errors together

- The decision procedure was designed to control α.
- We calculated β afterwards, rather than using β to design it.
- At fixed n, increasing α changes some negatives into positives, thus reducing false negatives (β) while increasing false positives.
- Likewise, decreasing α increases β.
- By increasing n, we can decrease β without increasing α. Increasing n results in a narrower power curve (previous slide).
- Goal: Find n to detect $p = .51$ with $\alpha = 0.05$.

Choosing n to control Type I and II errors together

Goal: Find n to detect $p = .51$ with $\alpha = 0.05$

General format of hypotheses for p in a binomial distribution

$$H_0\colon p = p_0$$

vs. one of these for H_1:

$$H_1\colon p > p_0$$
$$H_1\colon p < p_0$$
$$H_1\colon p \neq p_0$$

where p_0 is a specific value.

Our hypotheses

$$H_0\colon p = .5 \quad \text{vs.} \quad H_1\colon p > .5$$

Choosing n to control Type I and II errors together

Hypotheses

$$H_0\colon p = .5 \quad \text{vs.} \quad H_1\colon p > .5$$

- Flip the coin n times, and let x be the number of heads.
- Under the null hypothesis, $p_0 = .5$ so

$$z = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{x - .5n}{\sqrt{n(.5)(.5)}} = \frac{x - .5n}{\sqrt{n}/2}$$

- The z-score of $x = .51n$ is $z = \dfrac{.51n - .5n}{\sqrt{n}/2} = .02\sqrt{n}$
- We reject H_0 when $z \geqslant z_\alpha = z_{0.05} = 1.64$ (one-sided cutoff), so

$$.02\sqrt{n} \geqslant 1.64 \qquad \sqrt{n} \geqslant \frac{1.64}{.02} = 82 \qquad n \geqslant 82^2 = 6724$$

- Thus, if the test consists of $n = 6724$ flips, only $\approx 5\%$ of such tests on a fair coin would give $\geqslant 51\%$ heads.
- Increasing n further reduces the fraction α of tests giving $\geqslant 51\%$ heads with a fair coin.

Sign tests (nonparametric)

One-sample: Percentiles of a distribution

- Let X be a random variable. Is the 75th percentile of X equal to C?
- Get a sample $x_1, \ldots, x_n$.
- "Heads" is $x_i \leqslant C$, "tails" is $x_i > C$.
- Test

$$H_0 : \ p = .75 \qquad \text{vs.} \qquad H_1 : \ p \neq .75$$

- Of course this works for any percentile, not just the 75th.
- For the median (50th percentile) of a continuous symmetric distribution, the Wilcoxon signed rank test could also be used

Sign tests (nonparametric)

Two-sample (paired): Equality of distributions

- Assume X, Y are continuous distributions differing only by a shift, $X = Y + C$. Is $C = 0$?
- Get paired samples $(x_1, y_1), \ldots, (x_n, y_n)$.
- Do a hypothesis test for a fair coin, where $y_i - x_i > 0$ is heads and $y_i - x_i < 0$ is tails.
- To test $X = Y + 10$, check the sign of $y_i - x_i + 10$ instead.
- Wilcoxon on $y_i - x_i$ could be used for paired data and Mann-Whitney for unpaired data.