Introduction to statistical inference

Maarten Speekenbrink

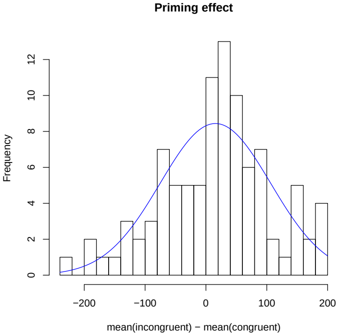
Experimental Psychology
University College London

Statistics lecture 2

Outline

1. Statistical inference
   - Introduction
   - Statistical decision making
   - Hypothesis testing
2. The F test
3. Decision errors and power
   - Estimating power
   - Improving power
4. Confidence intervals

A statistical model of retroactive priming



- We have data Y_i and model

$$Y_i \sim N(\mu, \sigma)$$

We do not know μ and σ

Some notation

Throughout the course, we will mainly use linear models. These have the general form

$$Y_i = \beta_0 + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \ldots + \beta_{p-1} \times X_{p-1,i} + \epsilon_i$$

where the β_j's (beta's) are parameters and the X_j's are predictor variables. The final term ϵ_i (epsilon) is an error term, the part of the data which is not predictable from the model. For the error, we usually assume

$$\epsilon_i \sim N(0, \sigma)$$

Some notation

Throughout the course, we will mainly use linear models. These have the general form

$$Y_i = \beta_0 + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \ldots + \beta_{p-1} \times X_{p-1,i} + \epsilon_i$$

where the β_j's (beta's) are parameters and the X_j's are predictor variables. The final term ϵ_i (epsilon) is an error term, the part of the data which is not predictable from the model. For the error, we usually assume

$$\epsilon_i \sim N(0, \sigma)$$

Note that the β_j's are generally unknown parameters of the population. To distinguish these from estimated parameters, we will use

$$Y_i = b_0 + b_1 \times X_{1i} + b_2 \times X_{2i} + \ldots + b_{p-1} \times X_{p-1,i} + e_i$$

for an estimated model.

A simple model

For consistency, we rewrite our population model

$$Y_i \sim N(\mu, \sigma)$$

in linear model form as

$$Y_i = \mu + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

and then

$$Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

(so $\beta_0 = \mu$)

Note that there are no predictor variables here. All we know is that the observations come from the same population.

Statistical inference

- Data is a sample from the “population”, a hypothetical construct defined as the collection of all possible observations (from all people at all times).
- We’d like to say something about the population (e.g., it’s mean), not about the sample.
- Because we’ve sampled from the population, we can use the data to infer properties of the population. This is *statistical inference*. Two (related) methods:
  - Estimation: given a sample (data), estimate population parameter and a range of (more or less) likely values
  - Testing: given the sample data, decide whether population parameter is equal to a hypothetical value
- Always need a model of the population

Is the retroactive priming effect real?

In Bem's (2011) study, the average retroactive priming effect was positive. Our model of retroactive priming can be stated in the following three equivalent ways:

$$\texttt{priming}_i \sim N(\mu, \sigma)$$

$$\texttt{priming}_i = \mu + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

$$Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

where the last version uses the standard GLM notation, so $Y_i = \texttt{priming}_i$ and $\beta_0 = \mu$.

Is the retroactive priming effect real?

In Bem's (2011) study, the average retroactive priming effect was positive. Our model of retroactive priming can be stated in the following three equivalent ways:

$$\begin{aligned}
\texttt{priming}_i &\sim N(\mu, \sigma) && \\
\texttt{priming}_i &= \mu + \epsilon_i && \epsilon_i \sim N(0, \sigma) \\
Y_i &= \beta_0 + \epsilon_i && \epsilon_i \sim N(0, \sigma)
\end{aligned}$$

where the last version uses the standard GLM notation, so $Y_i = \texttt{priming}_i$ and $\beta_0 = \mu$.

How likely is it that $\beta_0 \neq 0$?

Is the retroactive priming effect real?

In Bem's (2011) study, the average retroactive priming effect was positive. Our model of retroactive priming can be stated in the following three equivalent ways:

$$\begin{aligned} \texttt{priming}_i &\sim N(\mu, \sigma) & & \\ \texttt{priming}_i &= \mu + \epsilon_i & \epsilon_i &\sim N(0, \sigma) \\ Y_i &= \beta_0 + \epsilon_i & \epsilon_i &\sim N(0, \sigma) \end{aligned}$$

where the last version uses the standard GLM notation, so $Y_i = \texttt{priming}_i$ and $\beta_0 = \mu$.

How likely is it that $\beta_0 \neq 0$?

Two decisions: decide $\beta_0 = 0$ or decide $\beta_0 \neq 0$.

Decisions, decisions. . . (slight detour)

- Two possible states: “ill” and “well”
- Two decisions: “treat” and “send home”
- Each combination of state and decision has a consequence (utility)

		Patient	
		Ill	Well
Doctor	Treat	a	b
	Send home	c	d

- Based on symptoms, doctor can assign probabilities $p(\text{ill})$ and $p(\text{well}) = 1 - p(\text{ill})$
- Principle of maximum expected utility: choose the action with the highest expected utility.
  - decide to treat if

$$p(\text{ill}) \times a + p(\text{well}) \times b > p(\text{ill}) \times c + p(\text{well}) \times d$$

  - decide to send home otherwise

Decisions, decisions. . . (slight detour)

- Two possible states: “ill” and “well”
- Two decisions: “treat” and “send home”
- Each combination of state and decision has a consequence (utility)

		Patient	
		Ill	Well
Doctor	Treat	a	b
	Send home	c	d

- Based on symptoms, doctor can assign probabilities $p(\text{ill})$ and $p(\text{well}) = 1 - p(\text{ill})$
- Principle of maximum expected utility: choose the action with the highest expected utility.
  - decide to treat if

$$p(\text{ill}) \times a + p(\text{well}) \times b > p(\text{ill}) \times c + p(\text{well}) \times d$$

  - decide to send home otherwise

Decisions, decisions... (slight detour)

- Two possible states: “ill” and “well”
- Two decisions: “treat” and “send home”
- Each combination of state and decision has a consequence (utility)

		Patient	
		Ill	Well
Doctor	Treat	a	b
	Send home	c	d

- Based on symptoms, doctor can assign probabilities $p(\text{ill})$ and $p(\text{well}) = 1 - p(\text{ill})$
- Principle of maximum expected utility: choose the action with the highest expected utility.
  - decide to treat if

$$p(\text{ill}) \times a + p(\text{well}) \times b > p(\text{ill}) \times c + p(\text{well}) \times d$$

  - decide to send home otherwise

Too many decisions. . .

- Rather uncommon that there are only two possible states; number of possible states (e.g., parameter values) often infinite
- Often one properly specified hypothesis, but alternative unspecified. E.g.
  - H_0 (null-hypothesis): $\beta_0 = B_0$ (e.g., mean priming effect in population is $B_0 = 0$).
  - H_a (alternative): $\beta_0 \neq B_0$ (mean priming effect in population is different from B_0).
- Again two decisions: accept H_0 or accept H_a. What is the best decision?

Too many decisions. . .

- Rather uncommon that there are only two possible states; number of possible states (e.g., parameter values) often infinite
- Often one properly specified hypothesis, but alternative unspecified. E.g.
  - H_0 (null-hypothesis): $\beta_0 = B_0$ (e.g., mean priming effect in population is $B_0 = 0$).
  - H_a (alternative): $\beta_0 \neq B_0$ (mean priming effect in population is different from B_0).
- Again two decisions: accept H_0 or accept H_a. What is the best decision?

Too many decisions...

- Rather uncommon that there are only two possible states; number of possible states (e.g., parameter values) often infinite
- Often one properly specified hypothesis, but alternative unspecified. E.g.
  - H_0 (null-hypothesis): $\beta_0 = B_0$ (e.g., mean priming effect in population is $B_0 = 0$).
  - H_a (alternative): $\beta_0 \neq B_0$ (mean priming effect in population is different from B_0).
- Again two decisions: accept H_0 or accept H_a. What is the best decision?

Too many decisions. . .

- Rather uncommon that there are only two possible states; number of possible states (e.g., parameter values) often infinite
- Often one properly specified hypothesis, but alternative unspecified. E.g.
  - H_0 (null-hypothesis): $\beta_0 = B_0$ (e.g., mean priming effect in population is $B_0 = 0$).
  - H_a (alternative): $\beta_0 \neq B_0$ (mean priming effect in population is different from B_0).
- Again two decisions: accept H_0 or accept H_a. What is the best decision?

Too many decisions...

- Rather uncommon that there are only two possible states; number of possible states (e.g., parameter values) often infinite
- Often one properly specified hypothesis, but alternative unspecified. E.g.
  - H_0 (null-hypothesis): $\beta_0 = B_0$ (e.g., mean priming effect in population is $B_0 = 0$).
  - H_a (alternative): $\beta_0 \neq B_0$ (mean priming effect in population is different from B_0).

  (a capital B_j is used to denote an a priori value of β_j; usually $B_0 = 0$, but this does not have to be the case)
- Again two decisions: accept H_0 or accept H_a. What is the best decision?

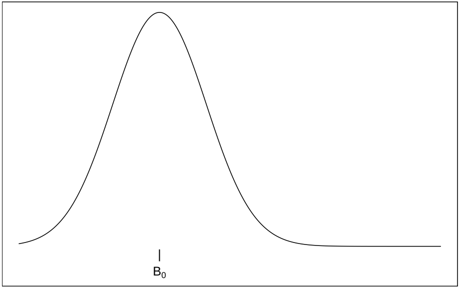
Too many decisions...

- Rather uncommon that there are only two possible states; number of possible states (e.g., parameter values) often infinite
- Often one properly specified hypothesis, but alternative unspecified. E.g.
  - H_0 (null-hypothesis): $\beta_0 = B_0$ (e.g., mean priming effect in population is $B_0 = 0$).
  - H_a (alternative): $\beta_0 \neq B_0$ (mean priming effect in population is different from B_0).

  (a capital B_j is used to denote an a priori value of β_j; usually $B_0 = 0$, but this does not have to be the case)
- Again two decisions: accept H_0 or accept H_a. What is the best decision?

Deciding between hypotheses

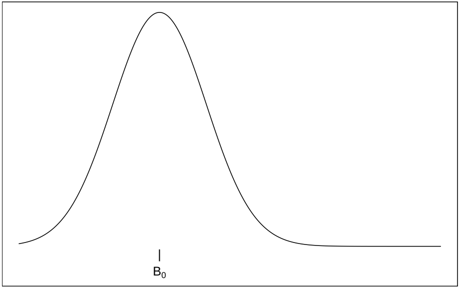
		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- In this simple model, the sample mean $\overline{Y}$ is the best estimator of β_0. Under H_0, we can work out the sampling distribution of $\overline{Y}$

Deciding between hypotheses

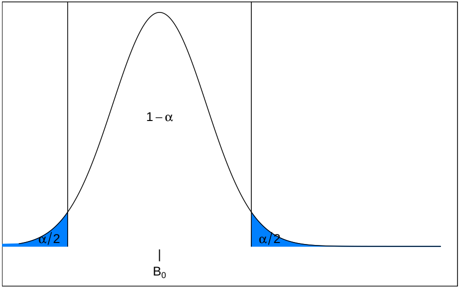
		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- If $\overline{Y}$ is unlikely when H_0 is true, we reject H_0
- We fix $p(\text{Type I error}|H_0) = \alpha$ by finding critical value(s). This significance level α is conventionally set to $\alpha = .05$

Deciding between hypotheses

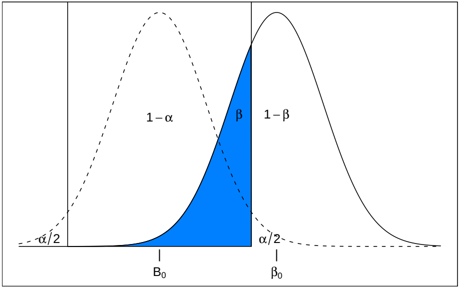
		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- If $\overline{Y}$ is unlikely when H_0 is true, we reject H_0
- We fix p(Type I error$|H_0) = \alpha$ by finding critical value(s). This significance level α is conventionally set to $\alpha = .05$

Deciding between hypotheses

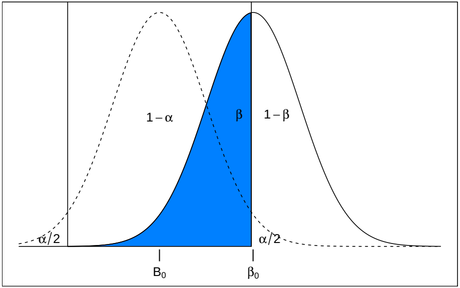
		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- Without knowing β_0, cannot determine $p(\text{Type II error}|H_a) = \beta$

Deciding between hypotheses

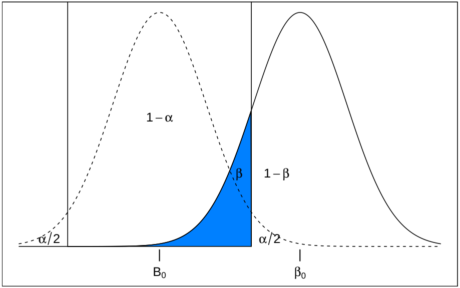
		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- Without knowing β_0, cannot determine $p(\text{Type II error}|H_a) = \beta$

Deciding between hypotheses

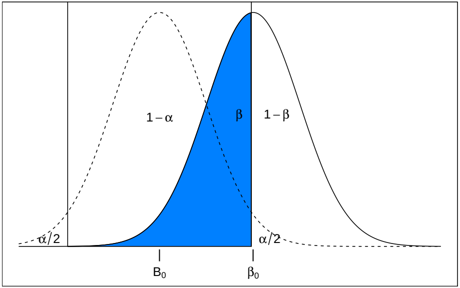
		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- Without knowing β_0, cannot determine $p(\text{Type II error}|H_a) = \beta$

Deciding between hypotheses

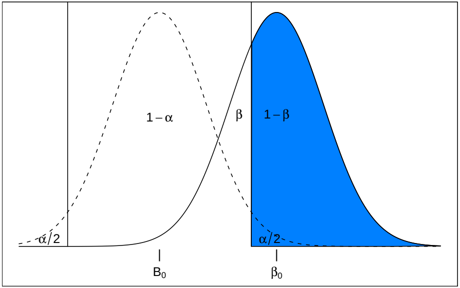
		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- Without knowing β_0, cannot determine $p(\text{Type II error}|H_a) = \beta$

Deciding between hypotheses

		population	
		H_0: $\beta_0 = B_0$	H_a: $\beta_0 \neq B_0$
decision	accept H_0	correct	type II error
	reject H_0	type I error	correct



- Without knowing β_0, cannot determine $p(\text{Type II error}|H_a) = \beta$
- The *power* of a test, $p(\text{reject } H_0|H_a) = 1 - \beta$, is thus also unknown

Null hypothesis significance testing

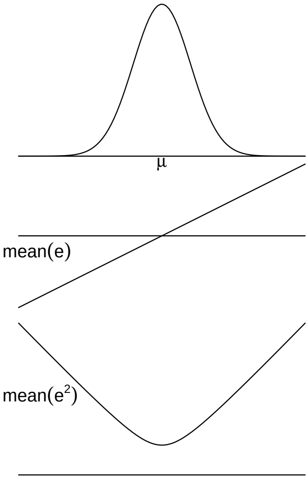
- Essentially: focus on a single hypothesis (H_0)
- Work out sampling distribution of a statistic, assuming H_0 is true
- If the value of the statistic in the data is unlikely given that H_0 is true, reject H_0
- Fix the probability of falsely rejecting H_0 to α, usually $\alpha = .05$
- The procedure is entirely based on keeping a specific error rate (Type I errors) within a certain bound. Without additional knowledge, we know nothing about the other error rate (Type II errors).

To estimate or not to estimate...

- If $\beta_0 = 0$, we should predict (future) priming effects as 0
- If $\beta_0 \neq 0$, we'd probably be better off using an estimated β_0 for future predictions
- Using $\hat{\beta}_0 = b_0 = \overline{Y}$ minimizes Sum of Squared Error in the sample
- Using $\beta_0 = \mu$ minimizes the Sum of Squared Error in the population
- $\overline{Y}$ is unlikely to be (exactly) equal to μ, so predictions $\hat{Y} = \overline{Y}$ are unlikely to be best for new samples
- If we have a prior idea about the value of β_0, would that lead to better predictions for future data?

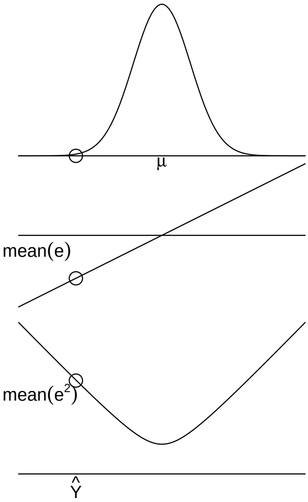
To estimate or not to estimate. . .

- If $\beta_0 = 0$, we should predict (future) priming effects as 0
- If $\beta_0 \neq 0$, we'd probably be better off using an estimated β_0 for future predictions
- Using $\hat{\beta}_0 = b_0 = \overline{Y}$ minimizes Sum of Squared Error in the sample
- Using $\beta_0 = \mu$ minimizes the Sum of Squared Error in the population
- $\overline{Y}$ is unlikely to be (exactly) equal to μ, so predictions $\hat{Y} = \overline{Y}$ are unlikely to be best for new samples
- If we have a prior idea about the value of β_0, would that lead to better predictions for future data?



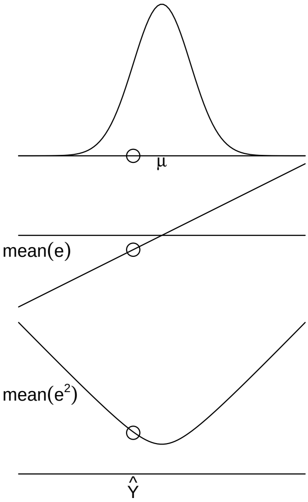
To estimate or not to estimate. . .

- If $\beta_0 = 0$, we should predict (future) priming effects as 0
- If $\beta_0 \neq 0$, we'd probably be better off using an estimated β_0 for future predictions
- Using $\hat{\beta}_0 = b_0 = \overline{Y}$ minimizes Sum of Squared Error in the sample
- Using $\beta_0 = \mu$ minimizes the Sum of Squared Error in the population
- $\overline{Y}$ is unlikely to be (exactly) equal to μ, so predictions $\hat{Y} = \overline{Y}$ are unlikely to be best for new samples
- If we have a prior idea about the value of β_0, would that lead to better predictions for future data?



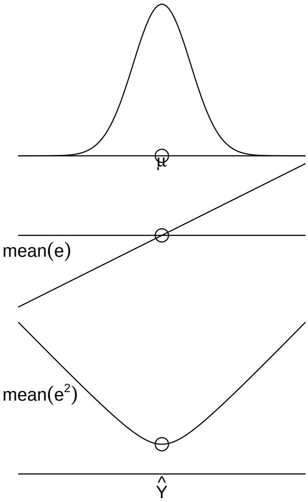
To estimate or not to estimate...

- If $\beta_0 = 0$, we should predict (future) priming effects as 0
- If $\beta_0 \neq 0$, we'd probably be better off using an estimated β_0 for future predictions
- Using $\hat{\beta}_0 = b_0 = \overline{Y}$ minimizes Sum of Squared Error in the sample
- Using $\beta_0 = \mu$ minimizes the Sum of Squared Error in the population
- $\overline{Y}$ is unlikely to be (exactly) equal to μ, so predictions $\hat{Y} = \overline{Y}$ are unlikely to be best for new samples
- If we have a prior idea about the value of β_0, would that lead to better predictions for future data?



To estimate or not to estimate...

- If $\beta_0 = 0$, we should predict (future) priming effects as 0
- If $\beta_0 \neq 0$, we'd probably be better off using an estimated β_0 for future predictions
- Using $\hat{\beta}_0 = b_0 = \overline{Y}$ minimizes Sum of Squared Error in the sample
- Using $\beta_0 = \mu$ minimizes the Sum of Squared Error in the population
- $\overline{Y}$ is unlikely to be (exactly) equal to μ, so predictions $\hat{Y} = \overline{Y}$ are unlikely to be best for new samples
- If we have a prior idea about the value of β_0, would that lead to better predictions for future data?



Back to the future

- Population model

$$Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0,\sigma)$$

- Two hypotheses:

$$H_0 : \beta_0 = 0$$
$$H_a : \beta_0 \neq 0$$

- Two models:

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0,\sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0,\sigma)$$

Back to the future

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- MODEL C (compact) is contained in (nested under) MODEL A (augmented)
  - The value 0 in MODEL C is one of the many possible values of β_0 in MODEL A
- Should we use MODEL C or MODEL A?
  - Compare the prediction error of MODEL C with the prediction error of MODEL A
  - When β_0 is estimated to minimize sample error, the sample error of MODEL A can never be higher than the sample error of MODEL C
  - We shouldn't just pick the model with the lowest sample error; we want to minimize the error in the population!

Back to the future

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- MODEL C (compact) is contained in (nested under) MODEL A (augmented)
  - The value 0 in MODEL C is one of the many possible values of β_0 in MODEL A
- Should we use MODEL C or MODEL A?
  - Compare the prediction error of MODEL C with the prediction error of MODEL A
  - When β_0 is estimated to minimize sample error, the sample error of MODEL A can never be higher than the sample error of MODEL C
  - We shouldn't just pick the model with the lowest sample error; we want to minimize the error in the population!

Back to the future

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- MODEL C (compact) is contained in (nested under) MODEL A (augmented)
  - The value 0 in MODEL C is one of the many possible values of β_0 in MODEL A
- Should we use MODEL C or MODEL A?
  - Compare the prediction error of MODEL C with the prediction error of MODEL A
  - When β_0 is estimated to minimize sample error, the sample error of MODEL A can never be higher than the sample error of MODEL C
  - We shouldn't just pick the model with the lowest sample error; we want to minimize the error in the population!

Back to the future

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- MODEL C (compact) is contained in (nested under) MODEL A (augmented)
  - The value 0 in MODEL C is one of the many possible values of β_0 in MODEL A
- Should we use MODEL C or MODEL A?
  - Compare the prediction error of MODEL C with the prediction error of MODEL A
  - When β_0 is estimated to minimize sample error, the sample error of MODEL A can never be higher than the sample error of MODEL C
  - We shouldn't just pick the model with the lowest sample error; we want to minimize the error in the population!

Back to the future

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- MODEL C (compact) is contained in (nested under) MODEL A (augmented)
  - The value 0 in MODEL C is one of the many possible values of β_0 in MODEL A
- Should we use MODEL C or MODEL A?
  - Compare the prediction error of MODEL C with the prediction error of MODEL A
  - When β_0 is estimated to minimize sample error, the sample error of MODEL A can never be higher than the sample error of MODEL C
  - We shouldn't just pick the model with the lowest sample error; we want to minimize the error in the population!

What is a substantial reduction in error?

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- We know that
$$\text{SSE}(\text{C}) \geq \text{SSE}(\text{A})$$
(the Sum of Squared Errors of MODEL C is at least as large as that for MODEL A). But is the difference substantial?
- If MODEL C is correct, then
$$F = \frac{\text{SSE}(\text{C}) - \text{SSE}(\text{A})}{\text{SSE}(\text{A})/(n-1)} = \frac{\text{SSR}}{\text{MSE}(\text{A})}$$
follows an F distribution with parameters $\text{df}_1 = 1$ and $\text{df}_2 = n - 1$

What is a substantial reduction in error?

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- We know that

$$\text{SSE}(\text{C}) \geq \text{SSE}(\text{A})$$

(the Sum of Squared Errors of MODEL C is at least as large as that for MODEL A). But is the difference substantial?

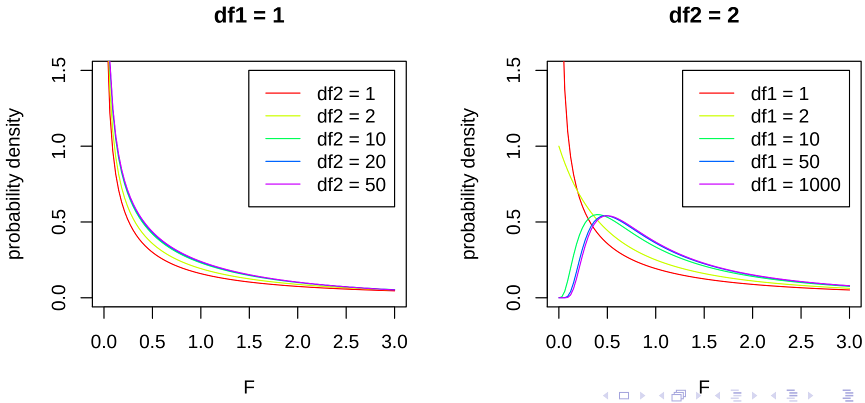
- If MODEL C is correct, then

$$F = \frac{\text{SSE}(\text{C}) - \text{SSE}(\text{A})}{\text{SSE}(\text{A})/(n-1)} = \frac{\text{SSR}}{\text{MSE}(\text{A})}$$

follows an F distribution with parameters $\text{df}_1 = 1$ and $\text{df}_2 = n-1$

The F distribution

- $F \geq 0$ (an F value is never negative!)
- Two parameters, df_1 and df_2
- Mean: $\frac{\text{df}_2}{\text{df}_2 - 2}$ (for $\text{df}_2 > 2$)



Back to the future 2

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

Back to the future 2

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

Estimated MODEL A:

$$\begin{aligned} Y_i &= b_0 + e_i \\ &= 15.58 + e_i \end{aligned}$$

Back to the future 2

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

Estimated MODEL A:

$$\begin{aligned} Y_i &= b_0 + e_i \\ &= 15.58 + e_i \end{aligned}$$

$$\text{SSE}(\text{A}) = 782{,}596.6$$
$$\text{SSE}(\text{C}) = 805{,}900.1$$

Back to the future 2

$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

Estimated MODEL A:

$$\begin{aligned} Y_i &= b_0 + e_i \\ &= 15.58 + e_i \end{aligned}$$

$$\text{SSE(A)} = 782,596.6$$
$$\text{SSE(C)} = 805,900.1$$

$$F = \frac{805,900.1 - 782,596.6}{782,596.6/(96-1)} = 2.829$$

Back to the future 2

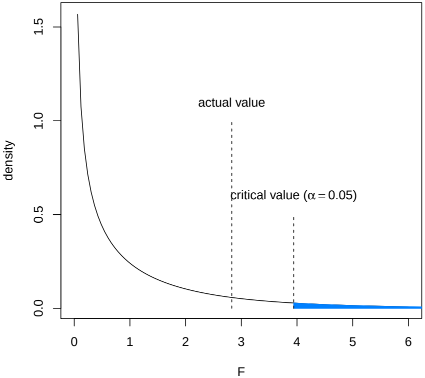
$$\text{MODEL C}: \quad Y_i = 0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$
$$\text{MODEL A}: \quad Y_i = \beta_0 + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

Estimated MODEL A:

$$\begin{aligned} Y_i &= b_0 + e_i \\ &= 15.58 + e_i \end{aligned}$$

$$\text{SSE(A)} = 782,596.6$$
$$\text{SSE(C)} = 805,900.1$$

$$F = \frac{805,900.1 - 782,596.6}{782,596.6/(96-1)} = 2.829$$



Proportional Reduction in Error

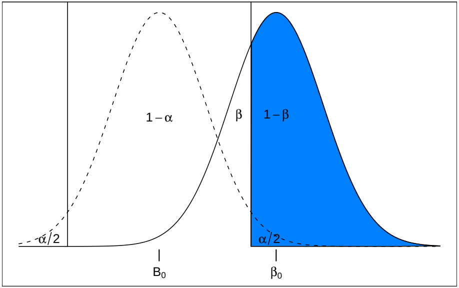
- While useful for hypothesis testing, the F statistic is not so easy to interpret
- A useful measure for effect size is the Proportional Reduction in Error (PRE)

$$\text{PRE} = \frac{\text{SSE}(\text{C}) - \text{SSE}(\text{A})}{\text{SSE}(\text{C})}$$

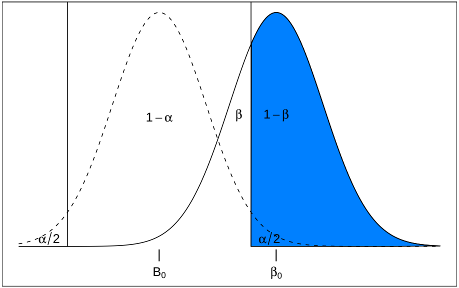
- The PRE has a value between 0 (useless) and 1 (extremely good)
- As with most things in statistics, we have to distinguish between the sample and population value of the PRE. The population value of the PRE is usually denoted as η^2 (eta-squared)
- For this data, PRE $= (805,900.1 - 782,596.6)/805,900.1 = .0289$

Power

- Remember: we have only fixed the probability of falsely rejecting the null hypothesis (Type I error)
- Because we don't know β_0, we don't know the probability of falsely accepting the null hypothesis (Type II error)
- If the difference between β_0 and B_0 is small, the probability of correctly rejecting the null hypothesis (power) may be very small indeed



Estimating power



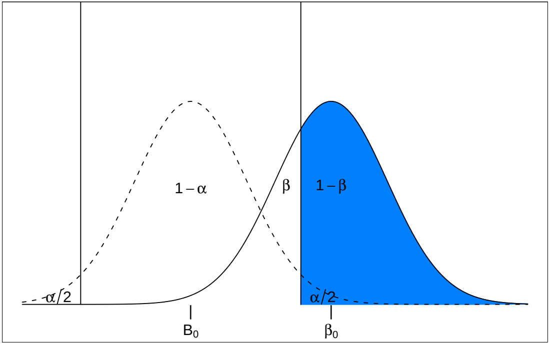
A priori power calculation:

- Computed using more or less sensible (a priori) values for PRE.

A posteriori power calculation:

- Some (including SPSS) estimate $1-\beta$ assuming estimated MODEL A is the true population model. They call this "observed power". This is not a great idea...

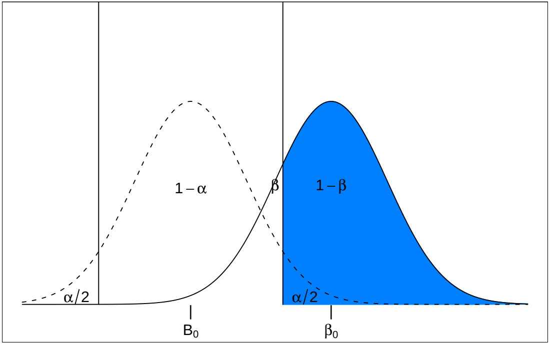
Improving power



Before

- Increasing α
- Decreasing error
- Increasing n

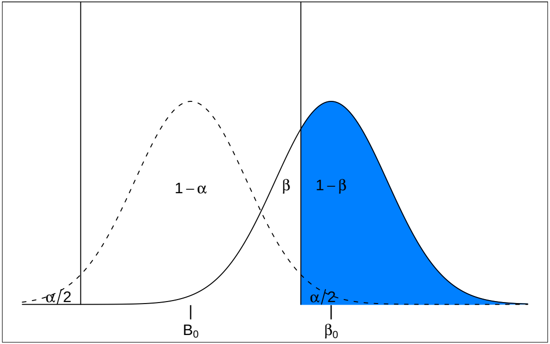
Improving power



After

- Increasing α
- Decreasing error
- Increasing n

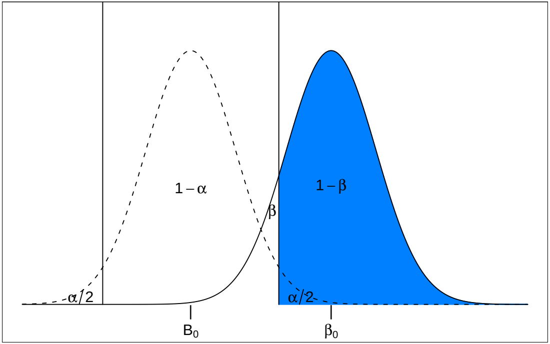
Improving power



Before

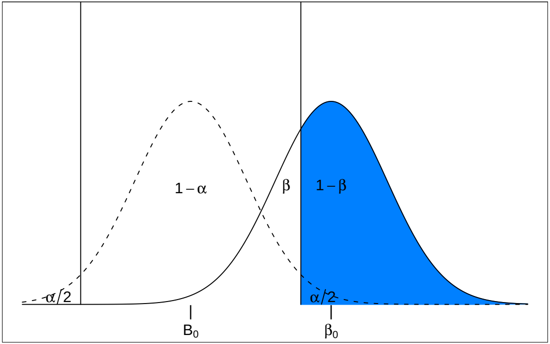
- Increasing α
- Decreasing error
- Increasing n

Improving power



- Increasing α
- Decreasing error
- Increasing n

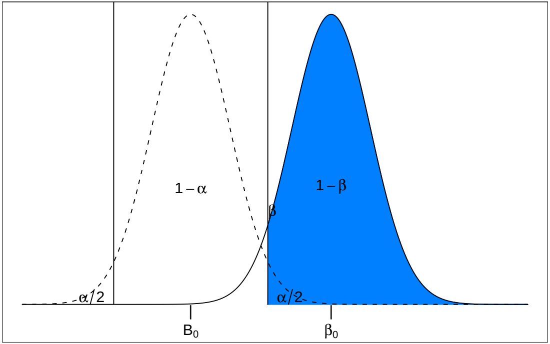
Improving power



Before

- Increasing α
- Decreasing error
- Increasing n

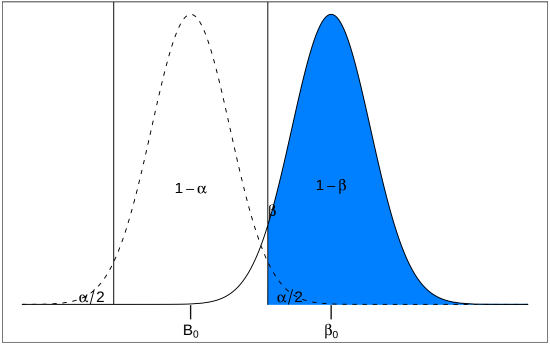
Improving power



After

- Increasing α
- Decreasing error
- Increasing n

Improving power



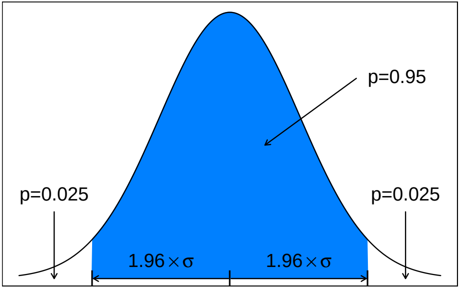
After

- Increasing α
- Decreasing error
- Increasing n

Decreasing error and increasing n both decrease the variance of the sampling distribution

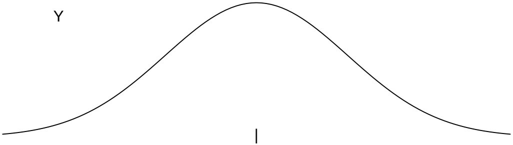
Confidence intervals

- An interval such that the probability that a value lies outside the interval has a predefined probability
  - e.g., $p = 0.95$ (or 95%) confidence interval
- For a Normal distribution, the width of a 95% confidence interval is approx. $3.92 \times \sigma$



- A confidence interval is computed from a sample, so has a *sampling distribution*

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

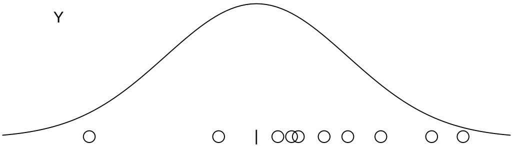
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

-
- Oops. . . population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

  Sampling distribution

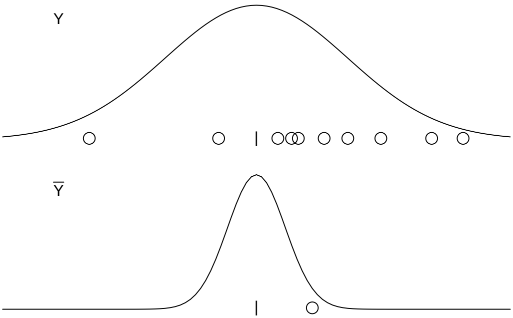
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

-
- Oops. . . population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

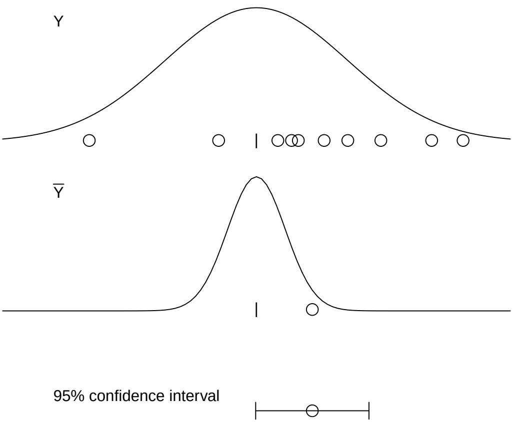
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

-
- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

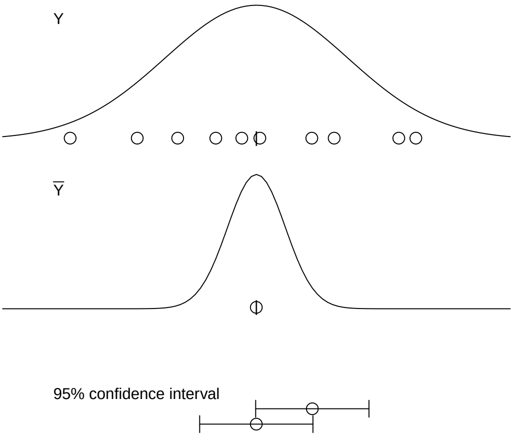
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

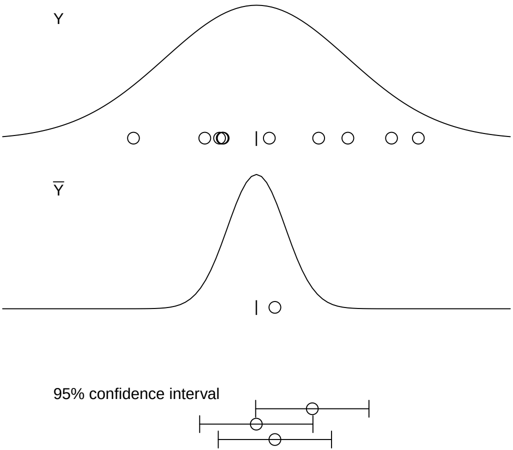
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

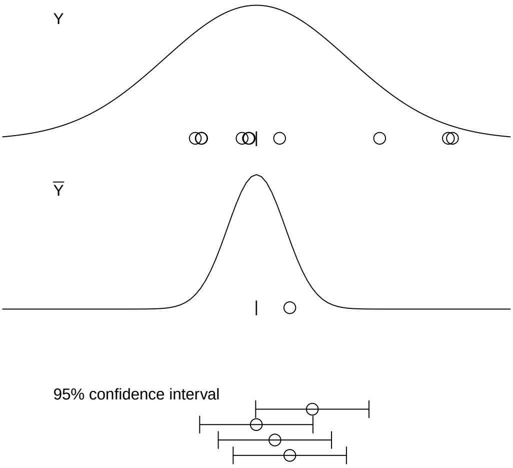
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops... population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

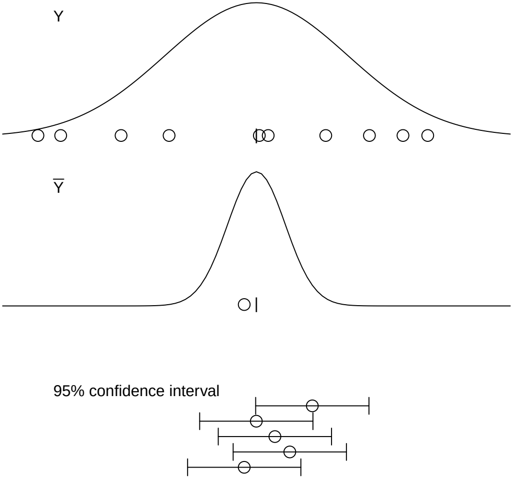
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

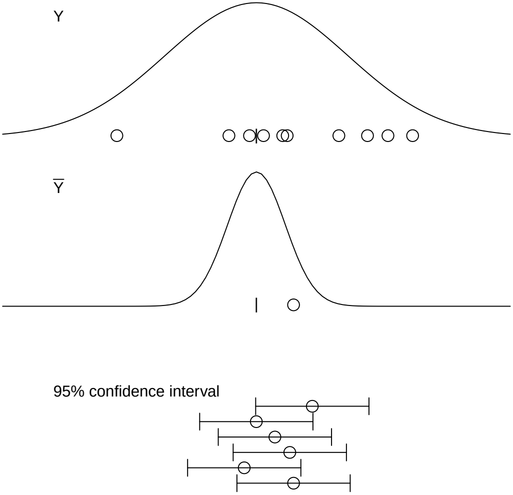
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

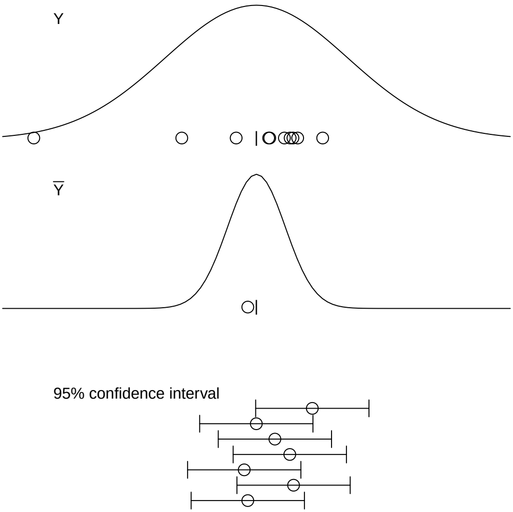
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again. . .
- Oops. . . population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

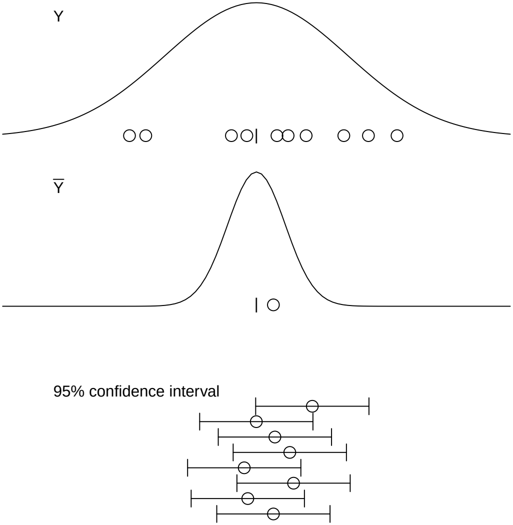
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

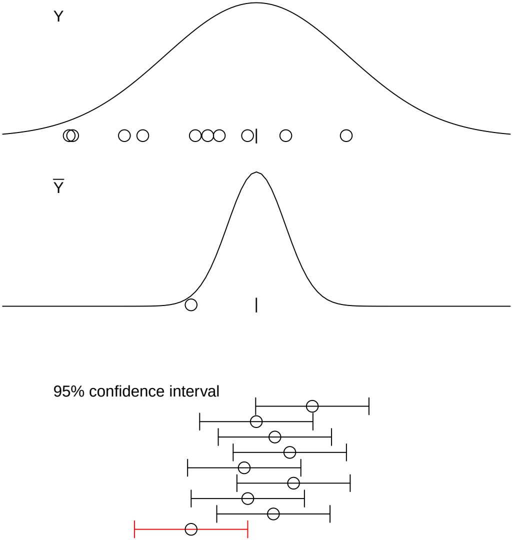
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

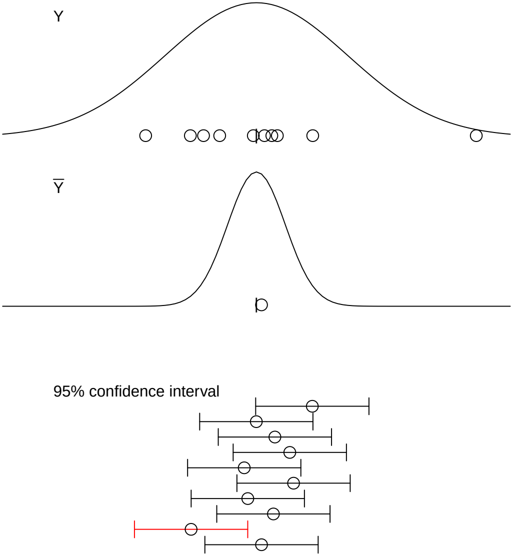
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

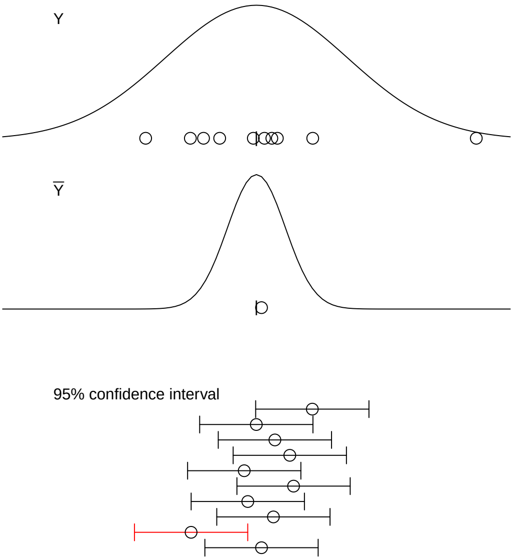
$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- And again...
- Oops... population mean outside interval

Confidence intervals (σ known)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

Sampling distribution

$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

- After infinite repetitions, 95% of the confidence intervals contain population mean

Oops. . . population mean

Computation

- Based on the sampling distribution of a parameter. In previous example, σ (population S.D.) was assumed known. If we have to estimate σ from sample data, we should take the resulting error into account in the confidence intervals.
- To estimate σ, we use the unbiased estimator of the population variance (note that $p = 1$ here)

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{n-p} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-p}$$

- Confidence interval for sample mean with estimated standard deviation $\hat{\sigma} = \sqrt{\text{MSE}}$:

$$\overline{Y} \pm t_{n-1;\alpha/2} \frac{\sqrt{\text{MSE}}}{\sqrt{n}}$$

$t_{n-1;\alpha/2}$ is critical value of t when df = $n-1$, and significance is α (e.g., $\alpha = .05$ for a 95% confidence interval).

Computation

- Based on the sampling distribution of a parameter. In previous example, σ (population S.D.) was assumed known. If we have to estimate σ from sample data, we should take the resulting error into account in the confidence intervals.
- To estimate σ, we use the unbiased estimator of the population variance (note that $p = 1$ here)

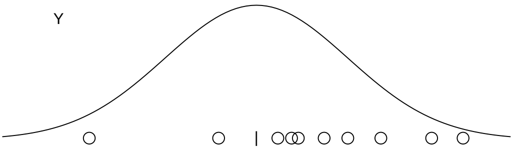
$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{n-p} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-p}$$

- Confidence interval for sample mean with estimated standard deviation $\hat{\sigma} = \sqrt{\text{MSE}}$:

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \frac{\sqrt{\text{MSE}}}{\sqrt{n}}$$

$F_{1,n-1;\alpha}$ is critical value of F when $\text{df}_1 = 1$, $\text{df}_2 = n-1$, and significance is α (e.g., $\alpha = .05$ for a 95% confidence interval).

Confidence intervals (σ unknown)



- Population distribution

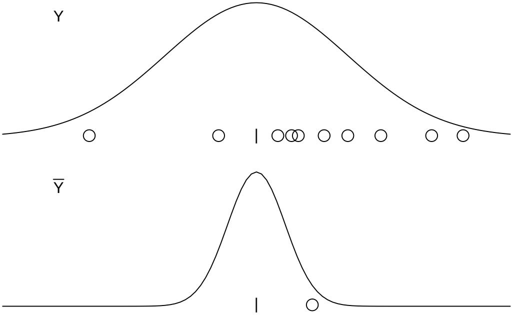
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- Oops. . . population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

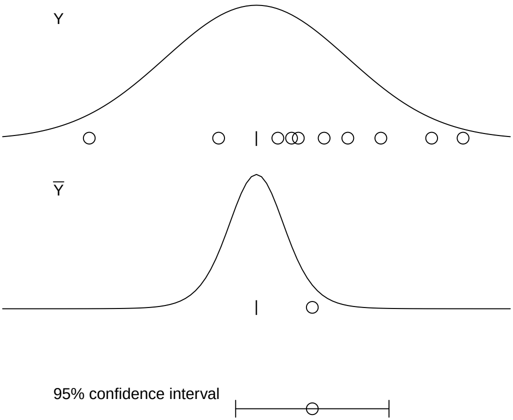
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

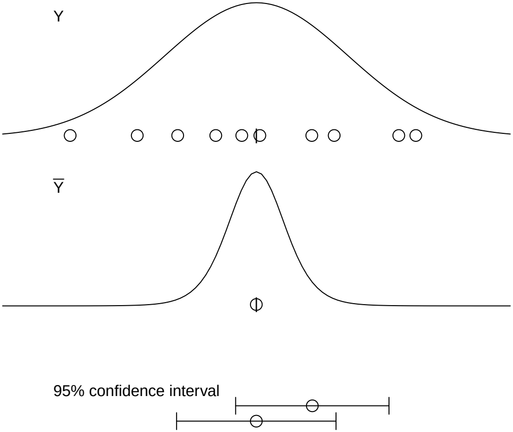
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

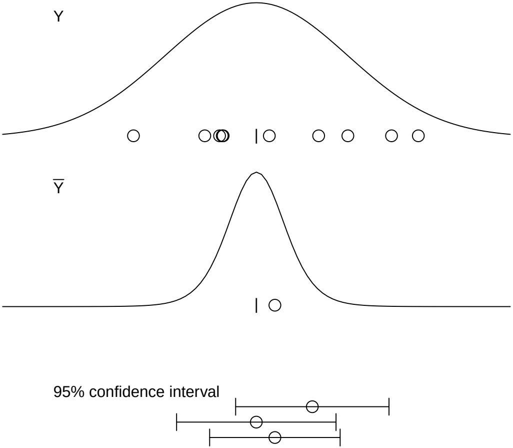
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

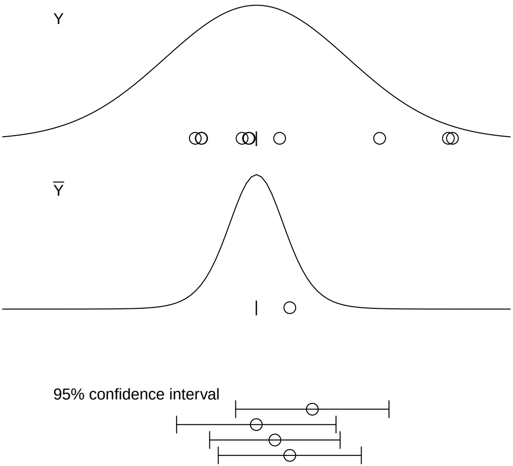
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

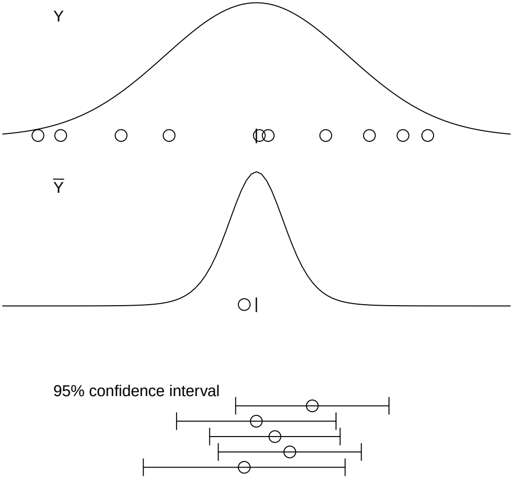
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

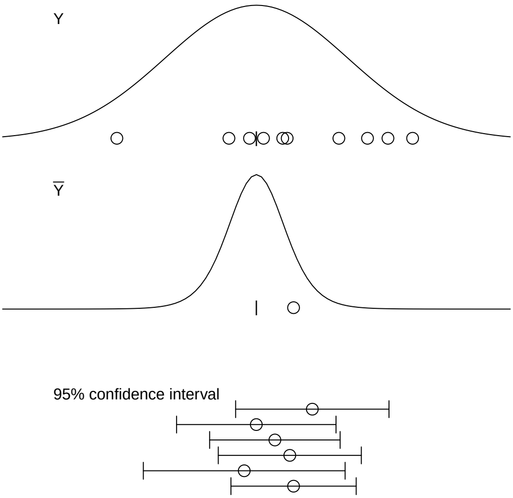
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops... population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

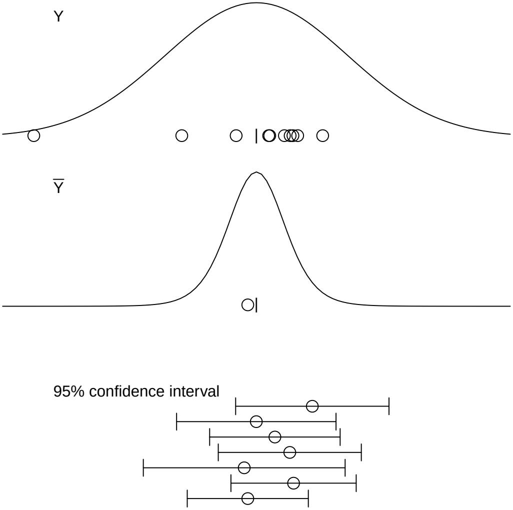
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again. . .
- Oops. . . population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

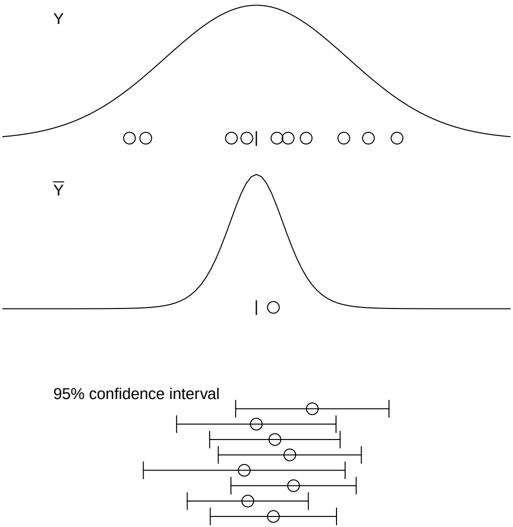
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops... population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

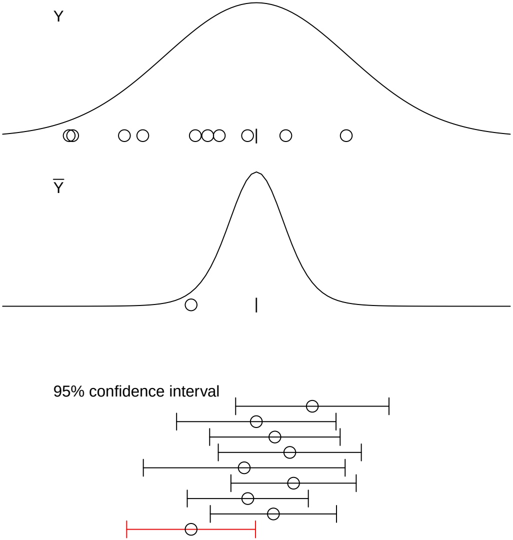
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

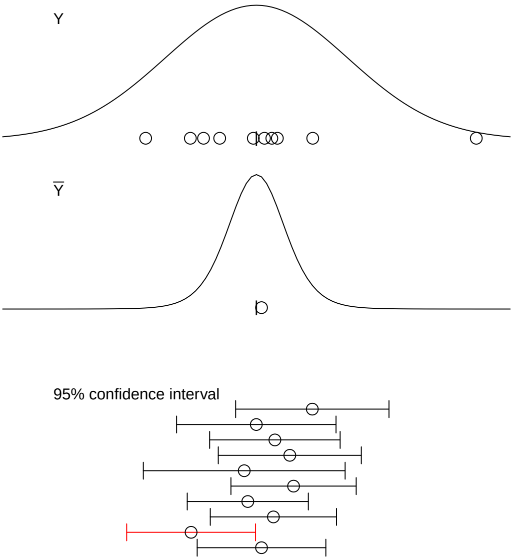
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

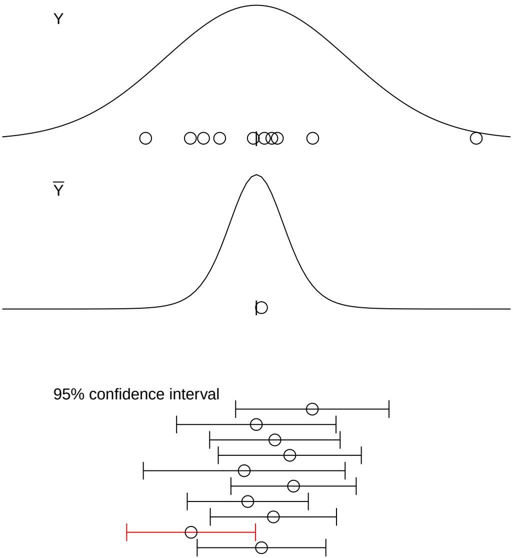
$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- And again...
- Oops...population mean outside interval

Confidence intervals (σ unknown)



- Population distribution

$$Y \sim N(\mu, \sigma)$$

- Take a sample
- Compute mean $\overline{Y}$
- Confidence interval

$$\overline{Y} \pm \sqrt{F_{1,n-1;\alpha}} \times \frac{\sqrt{MSE}}{\sqrt{n}}$$

- After infinite repetitions, 95% of the confidence intervals contain population mean
- Oops. . . population mean outside interval

Interpretation and usage

- Correct interpretation
  - if we take infinite samples from the population and compute the x% (or $1-\alpha$) confidence interval, then the true parameter lies within x% of those confidence intervals (which differ from sample to sample)
- Equivalent information to hypothesis test, e.g.
  - if B_0 does not lie in $1-\alpha$ confidence interval for β_0, this implies that MODEL C will be rejected in favour of MODEL A in a hypothesis test with a significance level of α.

Summary of key points

- Null hypothesis significance testing is about making decisions:
  - Should I accept or reject the null hypothesis H_0?
  - Based on a sampling distribution of a statistic of interest (e.g., sample mean), assuming H_0 is true.
  - Can't work out sampling distribution under the alternative hypothesis H_a, because it is usually too vague
- In the General Linear Model framework, hypothesis testing is a decision between estimating or fixing parameters. We base this decision on the comparison of two models:
  - A general (augmented) MODEL A in which parameters are estimated
  - A constrained (compact) MODEL C which fixes some of these parameters
  - If the error of the MODEL A is substantially smaller than the error of MODEL C, we decide to estimate and use MODEL A for our predictions
  - Error comparison based on the F-test
- Confidence intervals can provide same information as significance tests

Look ahead

The model comparison framework used today is general. Take MODEL A

$$Y_i = \beta_0 + \beta_1 \times X_{1i} + \ldots + \beta_{p-1} \times X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

and create MODEL C by fixing certain parameters β_j to a priori values B_j.

Decide whether to accept or reject these a priori values with an F-test:

$$\begin{aligned} F &= \frac{(\text{SSE}(\text{C}) - \text{SSE}(\text{A}))/(\text{PA} - \text{PC})}{\text{SSE}(\text{A})/(n - \text{PA})} \\ &= \frac{\text{MSR}}{\text{MSE}(\text{A})} \end{aligned}$$

where PA is the number of estimated parameters in MODEL A, and PC the number of estimated parameters in MODEL C

Further reading

This lecture:

- Judd, McClelland & Ryan, chapter 4

Next lecture:

- Judd, McClelland & Ryan, chapters 5 & 6