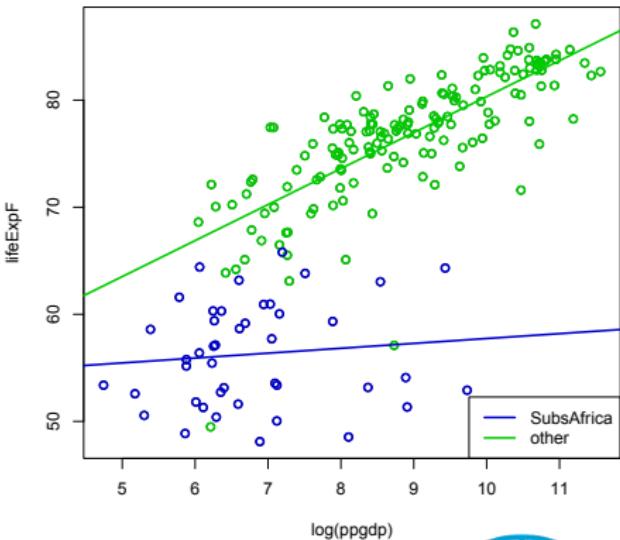
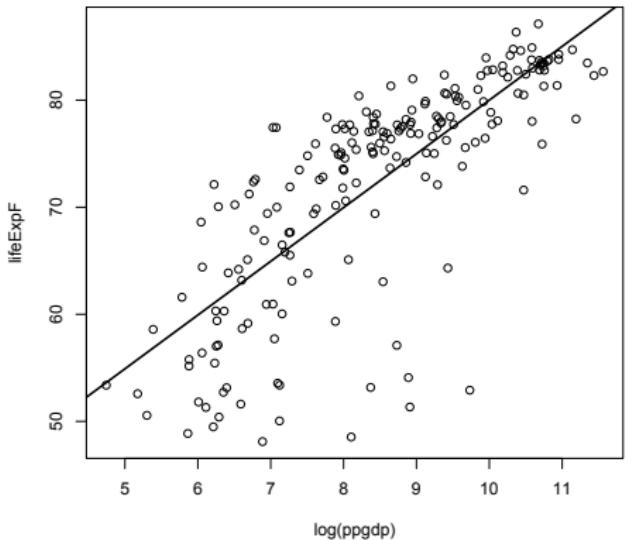


MODEL DIAGNOSTIC PLOTS



universität
wien

Gaussian linear Model: $Y \sim N(X\beta, \sigma^2 I_n)$



What if the model is too complex to visualize it like that? Can we still get a graphical representation of model fit?



MODEL DIAGNOSTIC PLOTS



universität
wien

Gaussian linear Model:

$$Y \sim N(X\beta, \sigma^2 I_n)$$

or, equivalently

$$Y_i = X_{i \cdot} \beta + u_i, \quad \text{with} \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \sim N(0, \sigma^2 I_n)$$

Recall: OLS

$$\hat{\beta} = \operatorname{argmin}_{b \in \mathbb{R}^p} \|Y - Xb\|_2^2 = \operatorname{argmin}_{b \in \mathbb{R}^p} \sum_{i=1}^n (Y_i - X_{i \cdot} b)^2$$

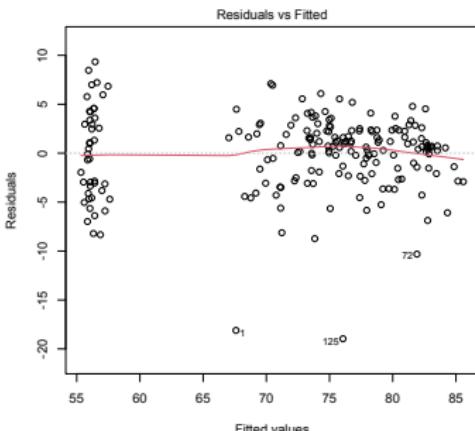
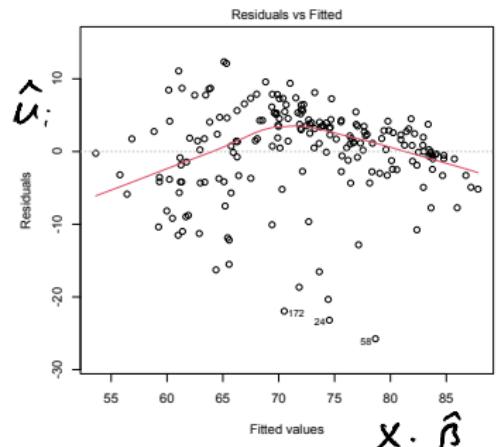
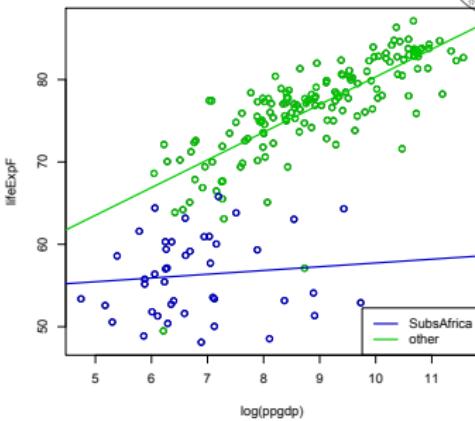
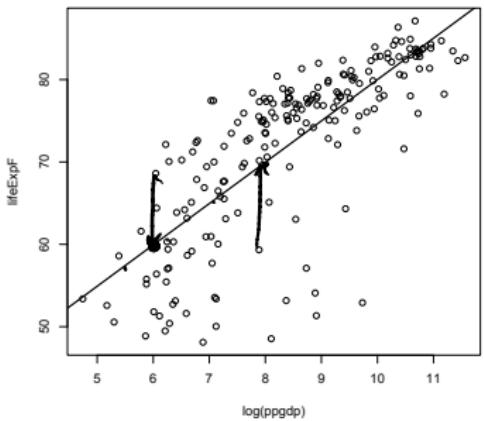
Idea: The *residuals* $\hat{u}_i = Y_i - X_{i \cdot} \hat{\beta}$ should be (approximately) iid Gaussian and evenly distributed around the regression line (independent of their 'location' on the regression line).

$\hat{Y}_i = X_{i \cdot} \hat{\beta}$ are the *fitted* or *predicted* values 'on' the regression line.

1.) plot \hat{Y}_i against \hat{u}_i

2.) check \hat{u}_i for Gaussianity

RESIDUAL PLOT

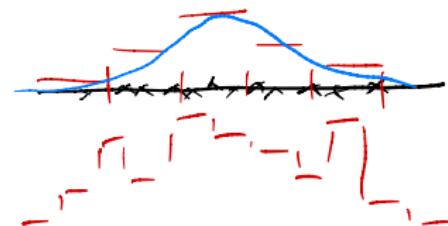


$1 = \text{Afghanistan}$
 $72 = \text{Greenland}$
 $125 = \text{Nauru}$

How do we check Gaussianity of the sample of residuals \hat{u}_i ,
 $i = 1, \dots, n$?

Construct a normal QQ-plot:

- ▶ Compute $p_i := \frac{\text{rank}(\hat{u}_i)}{n+1} \approx \frac{\text{rank}(\hat{u}_i)}{n}$.
- ▶ Compute $y_i := \Phi^{-1}(p_i)$.

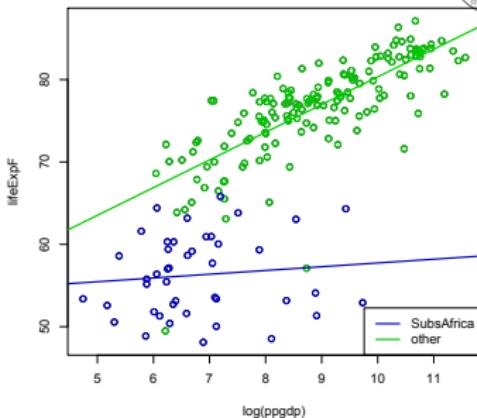
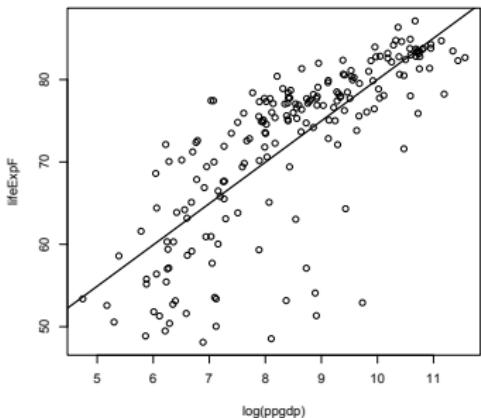


Plot the y_i against the **standardized residuals \tilde{u}_i** and draw the 45-degree line.

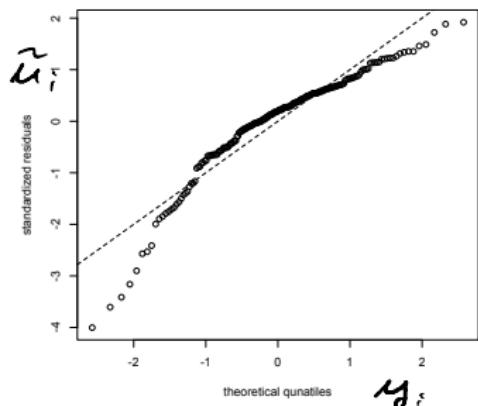
$$\tilde{u}_i = \frac{\hat{u}_i - \bar{\hat{u}}_n}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (\hat{u}_i - \bar{\hat{u}}_n)^2}}, \quad \bar{\hat{u}}_n = \frac{1}{n} \sum_{i=1}^n \hat{u}_i$$

$$y_i = \tilde{u}_i$$

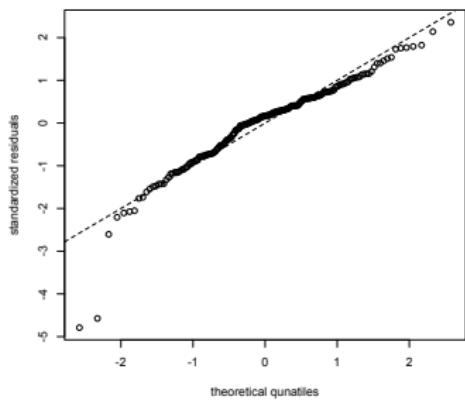
QQ PLOT



Normal QQ-plot

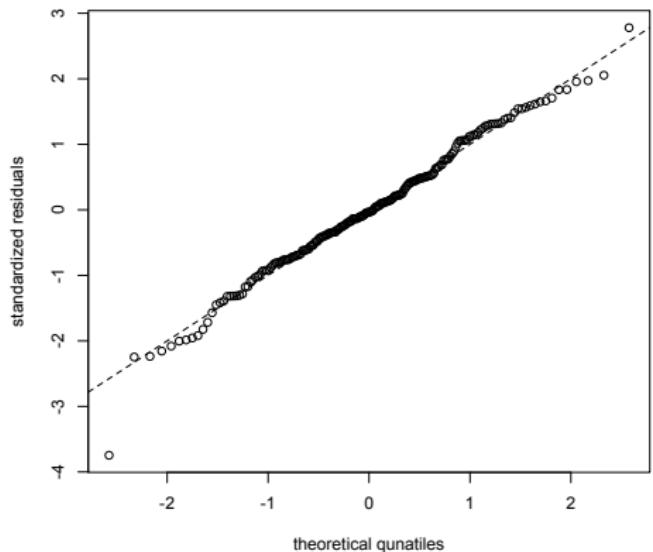


Normal QQ-plot

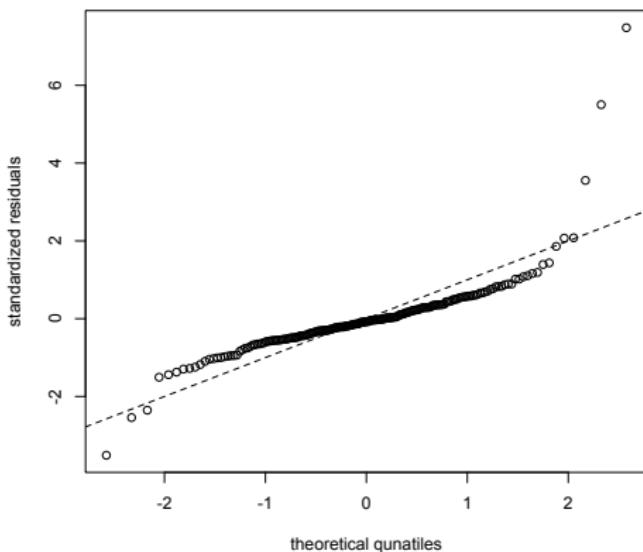


μ_f

Normal QQ-plot

 $N(0, 1)$ -data

Normal QQ-plot

 t_2 -data



FINDINGS ARE FALSE (IOANNIDIS, 2005)

Should we trust a study or interpretation that is based only on a single hypothesis test?

For simplicity: all tests have level α and power γ .

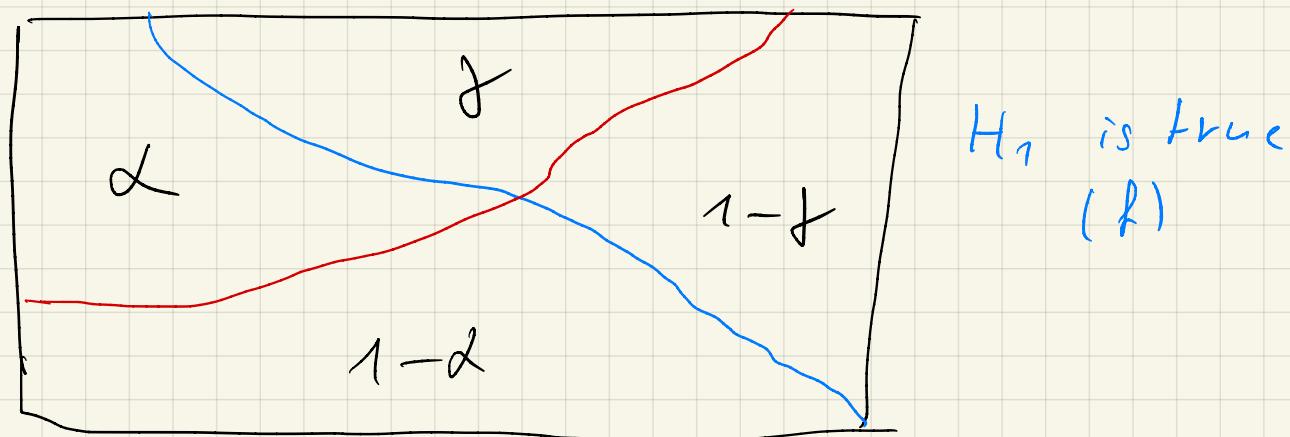
f is the fraction of all hypothesis tests conducted in a certain area of science (= the population) for which H_0 is actually false.

$$P(\text{true discovery} | \text{discovery}) = ?$$

$$P(\text{true discovery} \mid \text{discovery}) = \frac{P(\text{true discovery}, \text{discovery})}{P(\text{discovery})} = \frac{P(\text{true disc.})}{P(\text{disc.})}$$

H_0 rejected

all tests



H_0 is true

(1 - f)

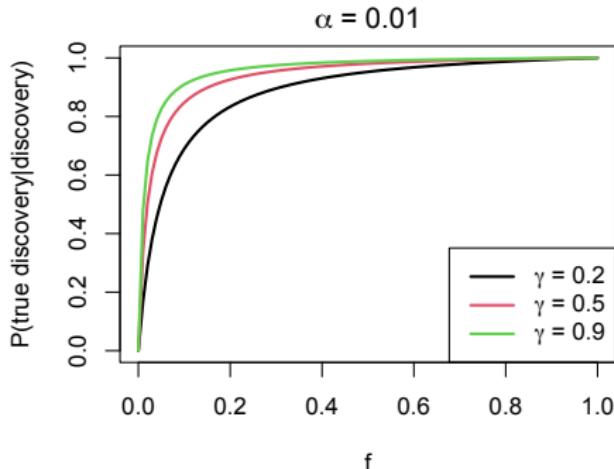
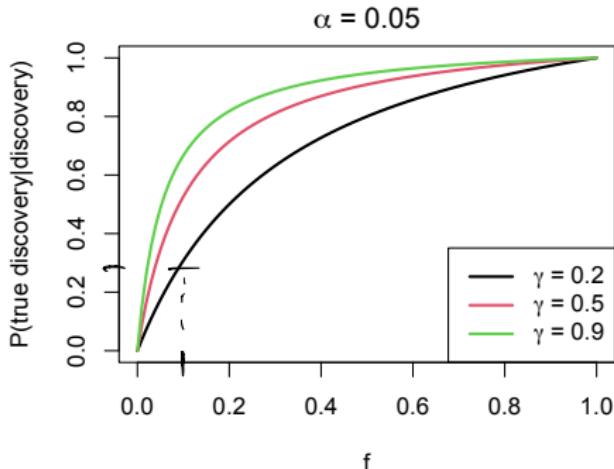
H_0 is accepted

H_1 is true
(f)

$$P(\text{true discovery}) = P(H_0 \text{ rejected}, H_1 \text{ true}) = f \cdot \beta$$

$$P(\text{discovery}) = f \cdot \beta + (1-f) \cdot \alpha$$

WHY MOST PUBLISHED RESEARCH FINDINGS ARE FALSE (IOANNIDIS, 2005)



- ▶ Ask: Is f in our case reasonably large?
Test hypotheses that are based on well established theories.
Just asking lots of questions will not give us more reliable answers.
- ▶ Try to reproduce your findings on independent follow-up studies!



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

NOVEMBER 5, 2020

VOL. 383 NO. 19

Remdesivir for the Treatment of Covid-19 — Final Report

J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luettkemeyer, S. Kline, D. Lopez de Castilla, R.W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, and H.C. Lane,
for the ACTT-1 Study Group Members*

ABSTRACT

BACKGROUND

Although several therapeutic agents have been evaluated for the treatment of coronavirus disease 2019 (Covid-19), no antiviral agents have yet been shown to be efficacious.

METHODS

We conducted a double-blind, randomized, placebo-controlled trial of intravenous remdesivir in adults who were hospitalized with Covid-19 and had evidence of lower respiratory tract infection. Patients were randomly assigned to receive either remdesivir (200 mg loading dose on day 1, followed by 100 mg daily for up to 9 additional days) or placebo for up to 10 days. The primary outcome was the time to recovery, defined by either discharge from the hospital or hospitalization for infection-control purposes only.

RESULTS

A total of 1062 patients underwent randomization (with 541 assigned to remdesivir and 521 to placebo). Those who received remdesivir had a median recovery time of 10 days (95% confidence interval [CI], 9 to 11), as compared with 15 days (95% CI, 13 to 18) among those who received placebo (rate ratio for recovery, 1.29; 95% CI, 1.12 to 1.49; $P<0.001$, by a log-rank test). In an analysis that used a proportional-odds model with an eight-category ordinal scale, the patients who received remdesivir were found to be more likely than those who received placebo to have clinical improvement at day 15 (odds ratio, 1.5; 95% CI, 1.2 to 1.9, after adjustment for actual disease severity). The Kaplan-Meier estimates of mortality were 6.7%

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Beigel at the National Institute of Allergy and Infectious Diseases, National Institutes of Health, 5601 Fishers Ln, Rm. 7E60, MSC 9826, Rockville, MD 20892-9826, or at jbeigel@niaid.nih.gov.

*A complete list of members of the ACTT-1 Study Group is provided in the Supplementary Appendix, available at NEJM.org.

A preliminary version of this article was published on May 22, 2020, at NEJM.org. This article was published on October 8, 2020, and updated on October 9, 2020, at NEJM.org.

N Engl J Med 2020;383:1813-26.

DOI: 10.1056/NEJMoa2007764

Copyright © 2020 Massachusetts Medical Society.

THE FILE DRAWER EFFECT



Can we use a published p -value just like that?

- ▶ Suppose we are doing a meta-analysis collecting 100 papers on clinical trials that investigated the efficacy of Remdesivir for Covid-19 treatment.
- ▶ Suppose they all applied sound statistical methodology.
⇒ Their p -values are valid!
- ▶ Suppose Remdesivir is really ineffective for treating Covid-19.
- ▶ In how many of our 100 papers do you expect to see a p -value of less than or equal to $\alpha_0 = 0.05$?

≈ 5

$$P_{H_0} (P \leq \alpha_0) \leq \alpha_0$$

THE FILE DRAWER EFFECT



Can we use a published p -value just like that?

- ▶ Suppose we are doing a meta-analysis collecting 100 papers on clinical trials that investigated the efficacy of Remdesivir for Covid-19 treatment.
- ▶ Suppose they all applied sound statistical methodology.
⇒ Their p -values are valid!
- ▶ Suppose Remdesivir is really ineffective for treating Covid-19.
- ▶ Suppose the peer review system only allows statistically significant results to be published.
- ▶ Suppose a result is considered significant if $p \leq \alpha_0 = 0.05$.
- ▶ In how many of our 100 papers do you expect to see a p -value of less than or equal to $\alpha_0 = 0.05$? *all of them*
- ▶ In how many of our 100 papers do you expect to see a p -value of less than or equal to $\alpha = 0.01$?



We are really looking at a subpopulation of all Remdesivir trials, namely those where $p \leq \alpha_0$.

$$P_{H_0}(p \leq \alpha | p \leq \alpha_0) = \alpha \quad ?$$

Recall Ex.2.2: A valid p-value of a simple null hypothesis is uniformly distributed under H_0 .

$$P_{H_0} (p \leq \alpha \mid p \leq \alpha_0) = \frac{P_{H_0}(p \leq \alpha, p \leq \alpha_0)}{P_{H_0}(p \leq \alpha_0)}$$

$$= \frac{\min(\alpha, \alpha_0)}{\alpha_0} \gg \alpha_0$$

$$\frac{0,01}{0,05} = 0,2$$

THE FILE DRAWER EFFECT



- ▶ Suppose a result is considered significant if $p \leq \alpha_0 = 0.05$.
- ▶ In how many of our 100 papers do you expect to see a p -value of less than or equal to $\alpha = 0.01$?

in approx. 20

Hence, the current peer review system may produce a lot more spurious discoveries than we would actually expect from controlling type-one error probabilities!

$$\begin{aligned} P_{\text{corr}} &:= \frac{P}{\alpha_0} \Rightarrow P_{H_0}(P_{\text{corr}} \leq \alpha | p \leq \alpha_0) \\ &= \frac{0.001}{0.05} \quad = P_{H_0}(p \leq \alpha \cdot \alpha_0 | p \leq \alpha_0) \\ &= \frac{\min(\alpha \cdot \alpha_0, \alpha_0)}{\alpha_0} \quad = \frac{\alpha \cdot \alpha_0}{\alpha_0} = \alpha. \end{aligned}$$

The misspecified Gaussian linear model

THE GAUSSIAN LINEAR MODEL



- ▶ $Y \sim N(X\beta, \sigma^2 I_n)$, $\beta \in \mathbb{R}^p$, $\sigma^2 \in (0, \infty)$.
- ▶ Low-dimensional case: $p < n$
- ▶ X is an $n \times p$ (non-random) design matrix with $\text{rank}(X) = p$ (e.g., analysis conditional on X)

These assumptions may be violated!

- ▶ That Y has Gaussian distribution can often be removed by appealing to the CLT approximation, i.e.,

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma} \sqrt{[(X'X)^{-1}]_k}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1),$$

- ▶ or by applying the bootstrap.
- ▶ But why would $\mathbb{E}[Y] = X\beta$ and $p < n$?

THE MISSPECIFIED GAUSSIAN LINEAR MODEL



- ▶ $Y \sim N(\mu, \sigma^2 I_n)$, $\mu \in \mathbb{R}^n$, $\sigma^2 \in (0, \infty)$.
- ▶ Low-dimensional case: $p < n$ (for now)
- ▶ X is an $n \times p$ (non-random) design matrix with $\text{rank}(X) = p$.
- ▶ For simplicity: $\sigma^2 = 1$.

We may still use OLS:

- ▶ $\hat{\beta} := \underset{b \in \mathbb{R}^p}{\text{argmin}} \|Y - Xb\|_2^2 = (X'X)^{-1}X'Y$
- ▶ $\hat{\beta} \sim N(\beta^*, \sigma^2(X'X)^{-1})$, where
- ▶ $\beta^* = \mathbb{E}[\hat{\beta}] = (X'X)^{-1}X'\mu = \underset{b \in \mathbb{R}^p}{\text{argmin}} \|\mu - Xb\|_2^2$.

“best approximation”

THE MISSPECIFIED GAUSSIAN LINEAR MODEL



$\beta = ?$ does not exist!

$$Y \sim N(\mu, I_n), \quad \mathbb{E}[Y] = \mu \approx X\beta^*$$

Suppose X_2 is 'gender' where 0 = female, 1 = male. If the i -th individual is female we have

$$\mathbb{E}[Y_i] = \mu_i \approx \beta_1^* + 0 + \sum_{k=3}^p \beta_k^* X_{ik},$$

whereas if the j -th individual is male, we have

$$\mathbb{E}[Y_j] = \mu_j \approx \beta_1^* + \beta_2^* + \sum_{k=3}^p \beta_k^* X_{jk}.$$

β_2^* is the additional income of men over women, given that all other variables are the same, in the best linear approximation to the expected income using the variables in X .

THE MISSPECIFIED GAUSSIAN LINEAR MODEL: STATISTICAL INFERENCE



universität
wien

$$Y \sim N(\mu, I_n), \quad \mathbb{E}[Y] = \mu \approx X\beta^* = \sum_{k=1}^p \beta_k^* X_{\cdot k}$$

Then

$$\frac{\hat{\beta}_k - \beta_k^*}{\sqrt{[(X'X)^{-1}]_k}} \sim N(0, 1),$$

where $[(X'X)^{-1}]_k$ is the k -th diagonal entry of $(X'X)^{-1}$.

Use this to construct tests or confidence intervals for β_k^* .

not for β !

THE MISSPECIFIED GAUSSIAN LINEAR MODEL



universität
wien

$$Y \sim N(\mu, I_n), \quad \mathbb{E}[Y] = \mu \approx X\beta^*$$

If we happen to choose the “correct” variables, i.e.,
 $\mu \in \text{span}(X)$, ...

$$\mu \in \text{span}(X) := \left\{ \begin{matrix} Xb \\ \in \mathbb{R}^n \end{matrix} : b \in \mathbb{R}^p \right\}$$

$$\Rightarrow \exists b_0 \in \mathbb{R}^p : \mu = Xb_0$$

$$\Rightarrow \beta^* = \underset{b \in \mathbb{R}^n}{\arg \min} \| \mu - Xb \|_2^2 = b_0$$

$$\Rightarrow Y \sim N(Xb_0, I_n)$$