# [STAFF] Solution for "Statistics for LSE" is missing some terms.

discussion posted 9 days ago by **DriftingWoods**

u^t on RHS of equation on line 2.

Extra X^T on LHS two times in each parentheses in equation on line 3.

Related to: Unit 6 Linear Regression:Lecture 20: Linear Regression 2 / 11. Significance Tests
This post is visible to everyone.

Add a Response

3 responses

ya_mukhin (Staff)
8 days ago

Thank you, @DriftingWoods, this has now been fixed.

Add a comment

? "Missing Parameter"??? 16
I am able to identify the corre...

Computing Kolmogorov-Smirnov tables in python 15
Community TA

? [STAFF] part (c) fixing a typo and the grader issue 4
In part (c), I entered **hat(sig...
Following

sandipan_dey
5 days ago

Since we have 5 features (genes), should not we add a $1^T$ vector (a column of 1s for the intercept) to have $p = 5 + 1 = 6$?

Add a comment

ya_mukhin (Staff)
5 days ago

Hi @sandipan_dey, in the model described in problem "Statistics for the LSE" the intercept term is known to be zero. So $p = 5$.

okay, thank you @ya_mukhin.

In a general setting, how does a model with an intercept term compare with a model without an intercept, trained on the same data? can we necessarily say / prove that one is better than the other, in terms of bias/variance (e.g., variance of the $\hat{\beta}$ estimates) / prediction error etc.?

posted 5 days ago by sandipan_dey

Hi @sandipan_dey, we know that the *fit* of the model with an intercept term will be better (smaller quadratic risk, prediction error) than the model without the intercept. This can be seen by thinking of the model without intercept as a *restricted* least squares estimator with the intercept coefficient set to zero. Provided we are in the small $p$, large $n$ regime, there should not be a noticeable change in the asymptotic distribution due to a single additional regressor. However, the variance of the error term is directly related to the fit of the model.

posted 4 days ago by ya_mukhin (Staff)

Thank you for the reply @ya_mukhin, so we have $Var(\epsilon) = \sigma^2 I_n$, with homoscedasticity assumption and also

$\hat{\beta}_{no\_intercept} = \beta^* + \mathcal{N}_p\left(0, \sigma^2 (X^T X)^{-1}\right)$ and

$\hat{\beta}_{with\_intercept} = \beta^* + \mathcal{N}_{p+1}\left(0, \sigma^2 (X_{w\_i}^T X_{w\_i})^{-1}\right)$, where $X_{w\_i} = [1, X]$,

so with intercept does only $Var(\hat{\beta})$ change and $Var(\epsilon)$ does not?

posted 4 days ago by sandipan_dey

Hi @sandipan_dey, when you say *in a general setting*, I think that there is no assumption of linearity of the conditional expectation and no assumption of homoskedasticity of the residual (no Gaussianity either). In general, conditional expectation $x \mapsto \mathbb{E}[Y|X=x]$ is some *measurable* function of $x$, does not need to be linear or even continuous. In this general setting, the *linear regression model* is the best linear approximation to the conditional expectation with respect to the quadratic risk.

posted 3 days ago by ya_mukhin (Staff)

Hi @ya_mukhin

1. Does the measurable function $x \mapsto \mathbb{E}[Y|X=x]$ maps from the same probability space $(\Omega, \mathcal{F}, \mathcal{P})$ (where $\mathcal{F}$ is a [Borel?] $\sigma$-algebra of the sample space $\Omega$) to the real numbers $\mathcal{R}$ as the random variable $X$ does? But wikipedia says about some sub-$\sigma$-algebras.
2. So we claim BLUE due to Gauss-Markov theorem right? will it be covered in the lectures? Does there a non-linear (e.g., kernelized) extension to the theorem exist?

Thank you

posted about 5 hours ago by sandipan_dey

@sandipan_dey, in probability theory we define the conditional expectation $E[Y|X]$ to be a random variable (defined on the same probability space as the random pair $(Y, X)$ that is measurable with respect to the sub-sigma algebra generated by $X$. For the purposes of statistics, we focus on the aspects of the random variables that can be "observed" i.e. used in computations. So we think of the conditional expectation with respect to $X$ as a function $x \mapsto E[Y|X=x]$ defined on the sample space $\mathbb{R}$ of $X$. The probabilistic definition of conditional expectation is just this function evaluated at $X$.

posted about 2 hours ago by ya_mukhin (Staff)

BLUE is a statement regarding optimality of the least squares estimator of the linear regression coefficients. When the error term is Gaussian, the LSE is also the MLE of these coefficients. The MLE is always efficient under regularity conditions. More generally, if the error terms are homoskedastic, then the LSE is efficient. With heteroskedastic error terms, there are different estimators that have a smaller asymptotic variance. The analysis of nonparametric kernel regression is very different from the analysis of least squares estimation of the linear regression and is outside the scope of this course.

posted about 2 hours ago by ya_mukhin (Staff)

Thank you very much @ya_mukhin for your patient reply to every question that I ask, really very helpful, could you please provide some relevant links on the analysis of nonparametric kernel regression? Thank you.

posted less than a minute ago by sandipan_dey

Add a comment

Showing all responses

## Add a response:

Preview

Submit