Purpose and methodology

Biological science research is often limited by the number of experimental units in a study. This is attributable to the cost of obtaining large numbers of experimental units and the time-scale required for the experimental units to produce data. For example, pigs that have been genetically modified to serve as organ donors for xenotransplantation require the zygote to be genetically modified, fertilized *in vitro*, and undergo a gestation period of around four months.[1,2] The cost and time for this process is thus very limiting and imposes experimental design limitations. The experimental design in studies ideally maximizes the detection of effects of treatment with the limited number of experimental units.

An important design question regarding experimental units is whether to distribute the experimental units amongst discrete treatment levels (ANOVA-design) or spread the experimental units along a continuous gradient of treatment levels (regression-design).[3] Here we compare the ability of these two experimental designs to detect a treatment effect by comparing their relative statistical power. This was done using R statistical software to estimate unknown parameters to explain potential relationships between the observed dependent variable and the controlled independent variable in two experimental data sets and in a *Monte Carlo* simulation.

Parameters were estimated using both least squares estimation (LSE) and maximum log-likelihood estimation (MLLE). In LSE, the unknown values of the parameters are estimated by finding numerical values for the parameters that minimize the sum of the squared errors in the model.[4] In MLLE, the unknown values of the parameters are estimated by finding numerical values for the parameters that maximize the probability of obtaining a specific data set, given the chosen probability distribution model.[5] All relevant scripts, simulation results, and associated documents can be found in the GitHub repository.

Antibiotics: ANOVA-design

In this analysis, we evaluated the effects of three different new antibiotics on growth of *E. coli* in lab cultures. Data was collected on bacterial growth after applying either no treatment (control), or one of three different antibiotics treatments. Given that the treatments can be thought of as categorical binary variables, an ANOVA-design was chosen to evaluate the effects. Using ANOVA, the means of growth of the groups were compared to the mean of growth for the control group that received no antibiotics. The results of the LSE and MLLE are displayed in Table 1. Both LSE and MLLE methods demonstrate similar parameter estimates and conclude a statistical difference in the control group relative to the treatment groups. In both cases, the complex model is a model fit with four estimates (one for each group), and the simple model is a model fit to no treatment at all. Resulting p-values of MLLE were calculated using a log-likelihood ratio (LLR) test statistic with three degrees of freedom on the chi-squared distribution (Table 1).

The expected growth of bacteria without antibiotics is 20.606 ± 2.701 divisions/hour; while antibiotic one is expected to decrease growth by 16.497 ± 3.819 divisions/hour, antibiotic two is expected to decrease growth by 3.085 ± 3.819 divisions/hour, and antibiotic three is expected to decrease growth by 12.269 ± 3.819 divisions/hour with 95% confidence (Table 1). Therefore, we can conclude that by applying antibiotic treatments one and three to *E. coli*, we will observe significantly reduced growth. Application of antibiotic treatment two was statistically ineffective as the 95% confidence interval of the true parameter associated with this group overlaps with a value of zero using both LSE and MLLE methods (Table 1). To further compare the differences between the groups, a t-test was performed comparing the control group to the treated groups. This is because ANOVA only tests for differences among the means, but does not tell specifically which means differ and to what extent. Given

that we are specifically trying to observe differences in group relative to only the control, the t-test was chosen. These results further support the statement that only antibiotic treatments one and three affect *E. coli* growth given that the p-values associated with those two treatments are significant at the $\alpha = 0.05$ level (Table 2). A visual representation of the differences between the control and treated groups is displayed in Figure 1.

Table 1. LSE and MLLE comparison of parameter estimates, residual standard error, and p-values.

Parameter Estimates	**MLLE**	**LSE**
Control	$\hat{\beta}_0 = 20.606 \pm 2.275$	$\hat{\beta}_0 = 20.605 \pm 2.700$
Antibiotic Treatment 1	$\hat{\beta}_1 = -16.497 \pm 3.217$	$\hat{\beta}_1 = -16.496 \pm 3.819$
Antibiotic Treatment 2	$\hat{\beta}_2 = -3.085772 \pm 3.217$	$\hat{\beta}_2 = -3.0853 \pm 3.819$
Antibiotic Treatment 3	$\hat{\beta}_3 = -12.269 \pm 3.217$	$\hat{\beta}_3 = -12.269 \pm 3.819$
Residual Standard Error	2.146	2.479
P-Value	2.965e-08	1.890e-06

Table 2. Resulting p-values of a t-test comparing control group to treated groups.

Treatment 1	**Treatment 2**	**Treatment 3**
$P = 5.632e\text{-}05$	$P = 0.0961$	$P = 0.005$

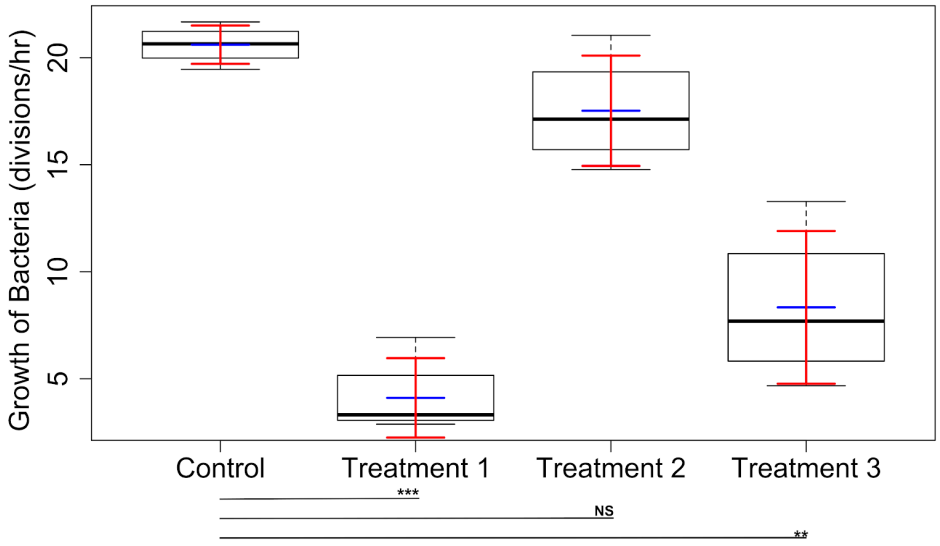

Figure 1: Boxplots representing the response of *E. coli* growth to different antibiotic treatments. For each categorical group, the boxplot shows the group's median growth (thick black line), average growth (blue line), quartile ranges (upper and lower bound black lines), and 95% confidence intervals of the means (red lines). Comparisons of each treated group were made to the control wherein treatment one had the greatest effect on *E. coli* growth, treatment three had a noticeable effect, and treatment two had no effect. (*** = $p < 0.001$, ** = $p < 0.01$, NS = Not Significant)

Sugar: regression-design

In this analysis, we evaluated the effects of sugar concentration on growth of *E. coli* in lab cultures. Data was collected for bacterial growth for varying concentrations of sugar in the culture media. Given that the concentration of sugar is a continuous variable, regression-design was chosen to evaluate the effects. The results of the LSE and MLLE are displayed in Table 3. Both LSE and MLLE methods demonstrate similar parameter estimates and conclude a statistically significant relationship between the growth of *E. coli* and the concentration of sugar in the growth media. In both cases, the complex model is a model fit with two estimates (one for the slope and one for the intercept of a regression line), and the simple model is a model fit only to the intercept. Resulting p-values of MLLE were calculated using a LLR test statistic with one degree of freedom on the chi-squared distribution (Table 3).

On average, for every one unit increase sugar concentration, the growth of *E. coli* will increase by 1.730 ± 0.297 divisions/hour with 95% confidence (Table 3). Therefore, we can conclude that there is a linear relationship between sugar concentration and *E. coli* growth. The coefficient of determination for this relationship was determined to be $R^2 = 0.9175$. This can be interpreted as 91.75% of the variability in the growth of *E. coli* can be explained by sugar concentration in the growth media. The approximated intercept parameter was deemed insignificant given that the 95% confidence interval of the true parameter value contains a zero value. This indicates that without any addition of sugar to the growth media, there is the possibility of no growth occurring. This result makes sense given that any organism will eventually die if deprived of nourishment. A visual representation of the linear relationship between the concentration of sugar in growth media and the measured growth of *E. coli* is displayed in Figure 2.

Table 3. LSE and MLLE comparison of parameter estimates, residual standard error, and p-values.

Parameter Estimates	**MLLE**	**LSE**
Intercept	$\widehat{\beta}_0 = -0.850 \pm 2.002$	$\widehat{\beta}_0 = -0.850 \pm 2.165$
Sugar Concentration Effects	$\widehat{\beta}_1 = 1.730 \pm 0.274$	$\widehat{\beta}_1 = 1.730 \pm 0.297$
Residual Standard Error	2.336	2.497
P-Value	2.638e-10	5.642e-09

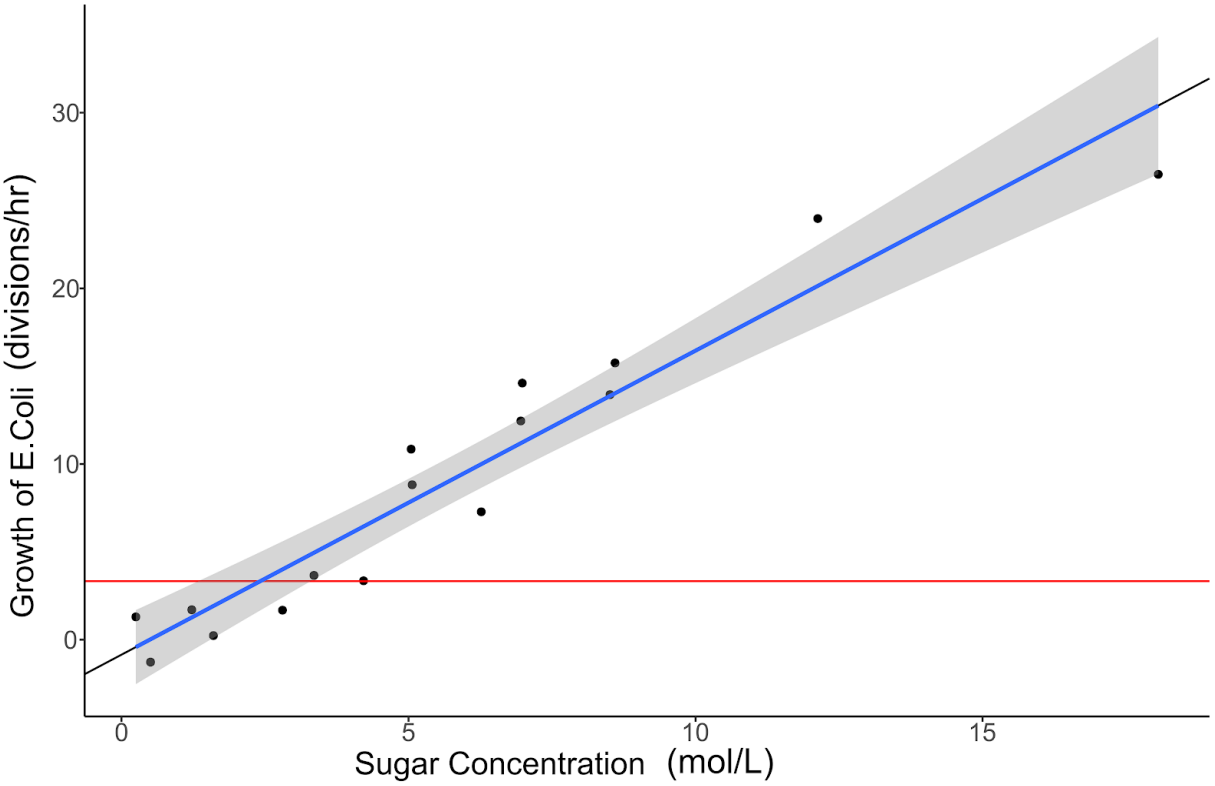

Figure 2: Linear relationship between sugar concentration in growth media and *E. coli* growth. For every one unit increase in sugar concentration, the growth of *E. coli* is on average expected to increase by 1.730 ± 0.297 divisions/hour. The blue line represents a regression line with a slope and intercept of the parameters optimized by MLLE. This regression line was compared to the simple model (red line) using the LLR statistic wherein the regression model better captures the relationship between sugar concentration and *E. coli* growth ($p < 0.001$, $R^2 = 0.9175$). Shaded in areas around the regression line represent the 95% confidence interval of the regression line.

Model specifications and assumptions

ANOVA: For ANOVA, the following assumptions are made[7]:

1. The experimental errors in the data are normally distributed
2. There exists equal variances between treatment groups (homoscedastic)
3. Samples are randomly selected and independent

Complex Model: $y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ \ldots \ + \beta_p \cdot x_p + \varepsilon$ where $\varepsilon \sim N(0,\sigma^2)$, p = number of groups, and the x_i's are binary response variables for the presence or absence of belonging to a particular group (In our ANOVA for the sugar data $p = 3$ and p = *number of levels* in the power analysis)

Simple Model: $y = \beta_0 + \varepsilon$ where $\varepsilon \sim N(0,\sigma^2)$

Regression: For linear regression, the following assumptions are made[6]:

1. The regression model is linear in the unknown parameters
2. The independent variable is fixed in repeated samples
3. The expected error in the model is zero on average
4. The variance-covariance matrix of the error is scalar
    a. Imposing that variance in the error of the model is equal across all data (homoscedastic)
    b. There exists no multicollinearity in the independent variable(s)

Complex Model: $y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ \ldots \ + \beta_p \cdot x_p + \varepsilon$ where $\varepsilon \sim N(0,\sigma^2)$ and p = number of continuous independent variables to be fit (In our linear regression experiments, $p = 1$)

Simple Model: $y \ = \ \beta_0 + \varepsilon$ where $\varepsilon \sim N(0,\sigma^2)$

Statistical tests for LSE and MLLE

LSE statistical testing in linear regression[8]: Statistical testing for LSE in linear regression tests if there does exist a linear relationship between the dependent and independent variables of a given data set. This results in the following hypothesis tests: $H_0 : \ \beta_1 = \ \beta_2 \ = \ \ldots \beta_{p-1} = 0$ and $H_a : \ at\ least\ one\ \beta_j \neq 0$ Thus, we are testing if the true values of the regression coefficients have some non-zero value. The associated test statistic is thus:

$$F \ = \ \frac{\left(\frac{RSS_{simple} - RSS_{complex}}{p_{complex} - p_{simple}}\right)}{\left(\frac{RSS_{simple}}{n - p_{complex}}\right)}$$

wherein: $RSS_{complex}$ is the residual sum of squares of the complex model, RSS_{simple} is the residual sum of squares of the simple model, $p_{complex}$ is the number of independent variables fit in the complex model, p_{simple} is the number of independent variables fit in the simple model, and n is the number of experimental units. This F statistic allows for calculation of a p-value due to the F distribution following the chi-squared distribution with $p_{complex}$ - p_{simple} degrees of freedom. At the $\alpha = 0.05$ level, if a calculated p-value is less than 0.05, then we can reject the null hypothesis and state that at least one parameter estimate does not equal zero in the fit model with 95% confidence. In other words, at least one parameter estimate demonstrates some linear relationship.

LSE statistical testing in ANOVA[9]: Statistical testing for LSE in ANOVA design tests if there does exist a difference in mean values of the dependent variable in separate categorical groups. This results in the following hypothesis tests: $H_0 : \ \mu_1 = \ \mu_2 \ = \ \ldots \ \mu_p$ and $H_a : \ at\ least\ one\ \mu_j \ differs$ Thus, we are testing if the true means of the independent variable are the same across all categorical groups. The associated test statistic is thus:

$$F = \frac{\sum_{i=1}^{K} n_i (\bar{Y}_{i\cdot} - \bar{Y})^2 / (K-1)}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2 / (N-K)}$$

wherein: the numerator represent the between-group variability, and the denominator represents the within-group variability; $\overline{Y}_i$ represents the sample mean in the i-th group, n_i is the number of observations in the i-th group, $\overline{Y}$ represents the overall mean of the dependent variable, K represents the number of groups, Y_{ij} is the j-th observation of the i-th group, and N is the overall sample size of the data set. This F statistic allows for calculation of a p-value due to the F distribution following the chi-squared distribution with $N-1$ degrees of freedom. At the $\alpha = 0.05$ level, if a calculated p-value is less than 0.05, then we can reject the null hypothesis and state that at least one mean does differ from the others with 95% confidence. Post-hoc tests can then be performed to find wherein the difference in group means lies. Though, given a control group as a reference with treated groups, a t-test can be

performed to compare means of groups to the control group, as done in the antibiotic ANOVA experiment as we are only interested in differences in means relative to the control.

MLLE statistical testing in linear regression and ANOVA[10]: Statistical testing for MLLE in both linear regression and ANOVA tests is used for determining if the likelihood function of a complex model with fit parameters is more plausible to explain the data distribution relative to a simple model with less fit parameters. This results in the following hypothesis tests: $H_0 : \theta \in \Theta_0$ wherein the parameters θ of the fit model exist in a subset Θ_0 of the total parameter space Θ, and $H_a : \theta \in \Theta_C$ wherein the parameters θ lie in the complement of Θ_0. In other words, the alternative hypothesis represents the complex model where the parameters of the fit have a defined value that help the likelihood function capture the data distribution, and the null hypothesis represents the simple model where the parameters do not capture the data distribution. Thus, under the null hypothesis, the reduced, simple model holds true while under the alternative hypothesis, the complex model holds true. The associated test statistic is thus the ratio between the log-likelihoods of the simple and complex model:[10]

$$\Lambda = 2\log\frac{\max_{\theta\in\Omega} f(X_1,\ldots,X_n|\theta)}{\max_{\theta\in\Theta_0} f(X_1,\ldots,X_n|\theta)} = 2(\mathrm{nll}_{\mathrm{simple}} - \mathrm{nll}_{\mathrm{complex}})$$

Wherein the test statistic simplifies to twice the difference of the negative log-likelihood of the simple model (nll_{simple}) and the negative log-likelihood of the complex model ($nll_{complex}$). This LLR test statistic allows for calculation of a p-value due to Wilk's theorem demonstrating that Λ tends to a chi-squared distribution with $p_{complex} - p_{simple}$ degrees of freedom for p representing the number of parameters estimated in each model.[11] At the $\alpha = 0.05$ level, if a calculated p-value is less than 0.05, then we can reject the null hypothesis and state that the more complex model is more plausible in explaining the distribution of the data.

Statistical Power Comparison of ANOVA and Linear Regression

After application of linear regression and ANOVA regression to two different data sets, a power analysis was conducted to compare the ability of both design methods to detect a theoretical treatment effect wherein the x variable is linearly related to the y variable with true parameter values of $\beta_1 = 0.4$ and $\beta_0 = 10$. Twenty-four experimental units were sequentially given x values from zero to fifty, and not randomly generated for each trial to preserve the assumptions above that state the independent variable must be fixed across repeated samples.[7] The dependent variable y then took values from the linear relationship give by the known parameter values of $\beta_1 = 0.4$ and $\beta_0 = 10$. Further, the y values were given random, normal error $\varepsilon \sim N(0,\sigma^2)$, in accordance with the complex model equations and assumptions above. To test ANOVA-design against regression-design, the average p-values across 10,000 simulations were compared for both MLLE and LSE methods across different observation levels and different standard deviations in the random error.

ANOVA-design split the independent variables into bins varying in length dependent upon the number of levels in the ANOVA-design. Means of the y values of the bins were then compared using ANOVA to test for differences in means as a method of detecting a theoretical treatment effect. Regression-design treated x as a continuous variable with a linear relationship with y wherein the slope of the linear relationship is used to detect the effects of a theoretical treatment. Ten-thousand simulations were run for two-, four-, and eight-level ANOVA design for each of the eight values of $\sigma = [1, 2, 4, 6, 8, 12, 16, 24]$. Thus, a total of 240,000 (3 levels · 8 sigma values · 10,000 simulations) simulations were run to compare the statistical power of ANOVA-design against regression-design wherein the

distributions of 960,000 p-values were compared from the combination of varying levels, sigmas, and parameter estimate methods (240,000 simulations · 2 model-designs · 2 parameter-estimation methods).

After running all simulations, observations were made across all levels and all sigma values to look for trends or differences in p-values across all methods. The p-value distributions for regression-design and ANOVA-design for both MLLE and LSE methods for eight-level ANOVA-design and σ = 24 were plotted (Figure 3). The regression-design p-value distribution for LSE methods had a right-skewed distribution with a mean p-value of 0.3177 ± 0.5667 (Figure 3A). The ANOVA-design p-value distribution for LSE methods had an unusual, right-skewed distribution wherein the first bin of p-values had a lower frequency than the subsequent bin with a mean p-value of 0.4221 ± 0.5631 (Figure 3B). The regression-design p-value distribution for MLLE methods had a right-skewed distribution with a mean p-value of 0.3013 ± 0.5675 (Figure 3C). The ANOVA-design p-value distribution had a right-skewed distribution with a mean p-value of 0.2908 ± 0.5326 (Figure 3D). All bounds of the p-values are calculated as the 95% confidence intervals. Across both design methods and both parameter estimation methods, mean p-values of all simulations are within 95% confidence intervals of one another. Thus, we can conclude with 95% confidence that similar p-values will be produced when using any combination of regression-design, ANOVA-design, LSE parameter estimation, or MLLE parameter estimation. Therefore, the statistical power of regression-design and ANOVA-design are the same, and both are viable options to choose when trying to detect treatment effects in an experiment.

To further test the statistical power of each method, the negative log base ten p-value distributions were observed for regression-design and ANOVA design for both MLLE and LSE methods using an ANOVA level of two and a sigma value of one. Regression-design p-value distribution for LSE methods under these conditions produced a normal distribution of values with a mean p-value of 2.738e-17 ± 5.367e-16 (Figure 4A). ANOVA-design p-value distribution for LSE methods also had a normal distribution with a mean p-value of 1.885e-07 ± 5.726e-07 (Figure 4B). Regression-design p-value distribution for MLLE methods produced p-values below machine epsilon; thus, the logarithmic value could not be calculated for them when comparing plots. To correct for this, a value of 1e-24 was added to each value to allow plotting of the distribution (Figure 4C). This was not done when calculating the average p-value nor the 95% confidence interval (3.552e-19 ± 1.793e-17). ANOVA-design p-value distributions for MLLE methods had a normal distribution with a mean p-value of 4.160e-08 ± 1.492e-07. (± values are calculated as 95% confidence intervals) From this, we can conclude that for any given, constant level of ANOVA-design and given, constant value of sigma, the p-values for ANOVA-design and regression-design will be normally distributed. This is a result of the method of adding random, normal error $\varepsilon \sim N(0,\sigma^2)$, to the y values. Given this, we observed no differences between using any combination of ANOVA-design, regression-design, LSE parameter estimation, or MLLE parameter estimation.

Because there were many non-significant p-values observed in Figure 3, we looked into where the non-significant p-values resulted from. To do this, regression-design and ANOVA-design average p-values and total counts of significant p-values were observed for all levels and all p-values and compared tablewise. Unsurprisingly, we found that regardless of the model that was being used, or the distribution of levels in x, that as σ was increased, the results of the simulations became less significant. As can be seen in Figure 5, the response of p-value shifts gradually from $p_{floor} << 0.0001$ for σ=1 to p_{ceil}=0.422138 for σ = 24. The seemingly critical σ at which the significance of the p-value becomes tenuous, is at σ = 8, which can be visualized by the gradient color scale wherein green indicates significant p-values, and yellow-red indicates increasingly insignificant p-values. Perhaps more interesting is the comparison of the results within the same σ across different experimental designs and their respective parameter estimation methods (observations down a column, Figure 5). The comparison across ANOVA-designs, where experimental units are divided into either 2 groups of 12, 4 groups of 6, or 8 groups of 3, did not affect the predictive power of the experiment given that average p-values and

total count of significant p-values are consistent across all levels. Given that all ANOVA models, regardless of the number of levels the experimental units are broken into, were statistically equivalent, the comparison between ANOVA-design and regression-design was observed for the different levels. This can be visualized again in Figure 5, with the comparison now being made between adjacent columns within the same σ value. Regardless of parameter estimation method (LSE or MLLE) both regression-design and ANOA-design produced similar counts of total significant p-values (Lin.Reg and adjacent ANOVA columns for any given σ and level, Figure 5).

Combining the results and observations, the choice of a regression-design in which *x* is characterized as a continuous variable or an ANOVA-design in which *x* is treated categorically in 2, 4, or 8 levels, did not significantly affect our ability to detect treatment effects (Figure 3, 5); these results were obtained by Monte Carlo simulated power analysis. Further, using LSE or MLLE as a method for parameter estimation did not affect our ability to detect treatment effects (Figure 3, 5). We conclude that the only impacting factor on p-values and statistical power is the amount of variance in the data set, as indicated by decreased significance as σ increased (Figure 5). Thus, when a scientist is designing experiments, both ANOVA-design and regression-design methods are comparable, and the deciding factors should be in how the scientist wishes to interpret their results and how much variance is implicit in the data set.

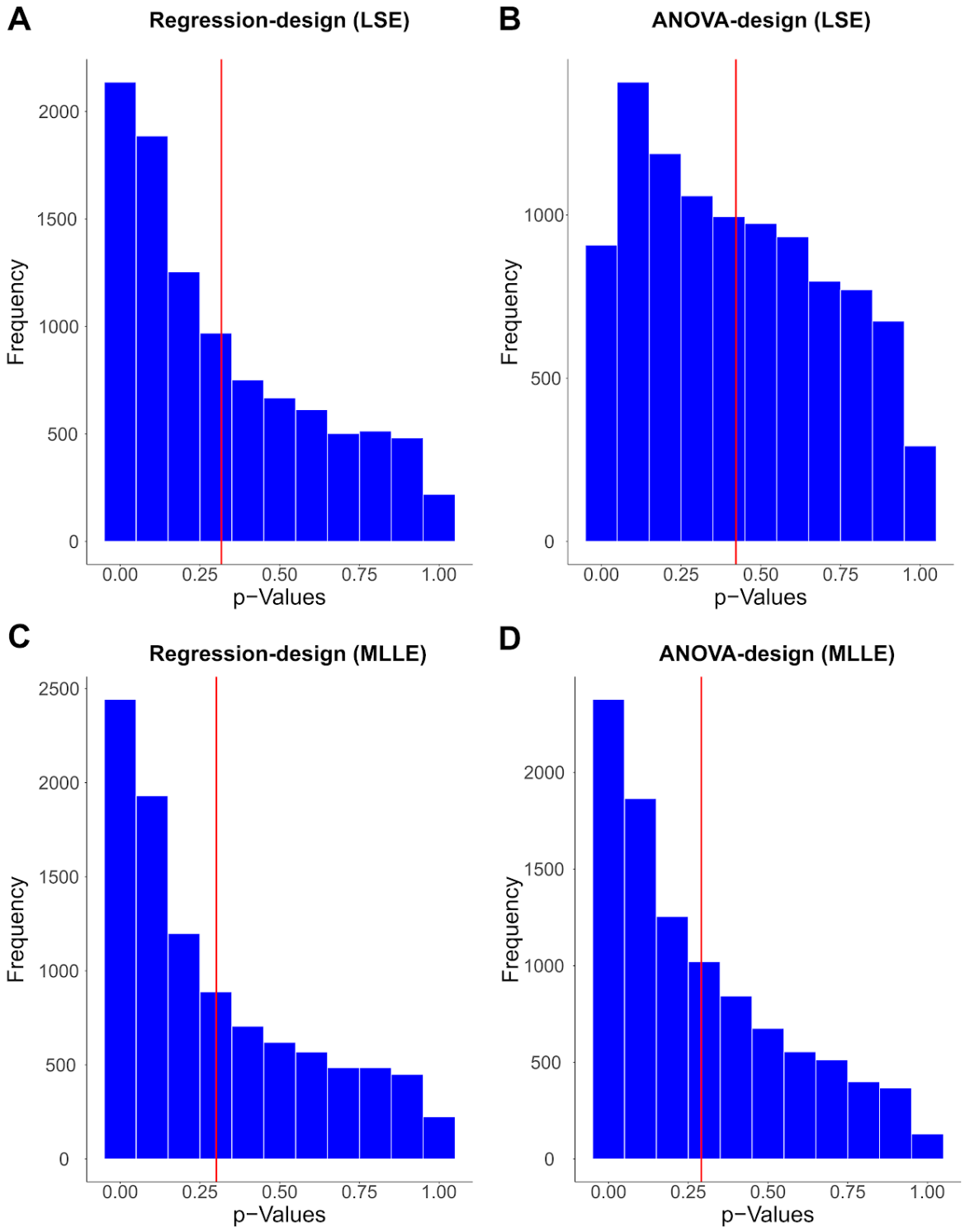

Figure 3: P-value distributions for regression-design and ANOVA-design for both MLLE and LSE methods for eight-level design and σ = 24. Across all simulations for all ANOVA-levels and all sigma values, the p-value distributions were plotted for regression-design and ANOVA-design for MLLE and LSE methods to compare statistical power. Average p-values across the simulations are indicated by the red line. A) Regression-design p-value distribution for LSE methods with a mean p-value of 0.31772203 ± 0.5667013. B) ANOVA-design p-value distribution for LSE methods with a mean p-value of 0.4221375 ± 0.5631884. C) Regression-design p-value distribution for MLLE methods with a mean p-value of 0.3013111 ± 0.567517. D) ANOVA-design p-value distributions for MLLE methods with a mean p-value of 0.2908227 ± 0.532657. Across all design methods and parameter estimation methods, p-values are within 95% confidence intervals of one another, indicating all are equally statistically powerful. (± values are calculated as 95% confidence intervals)

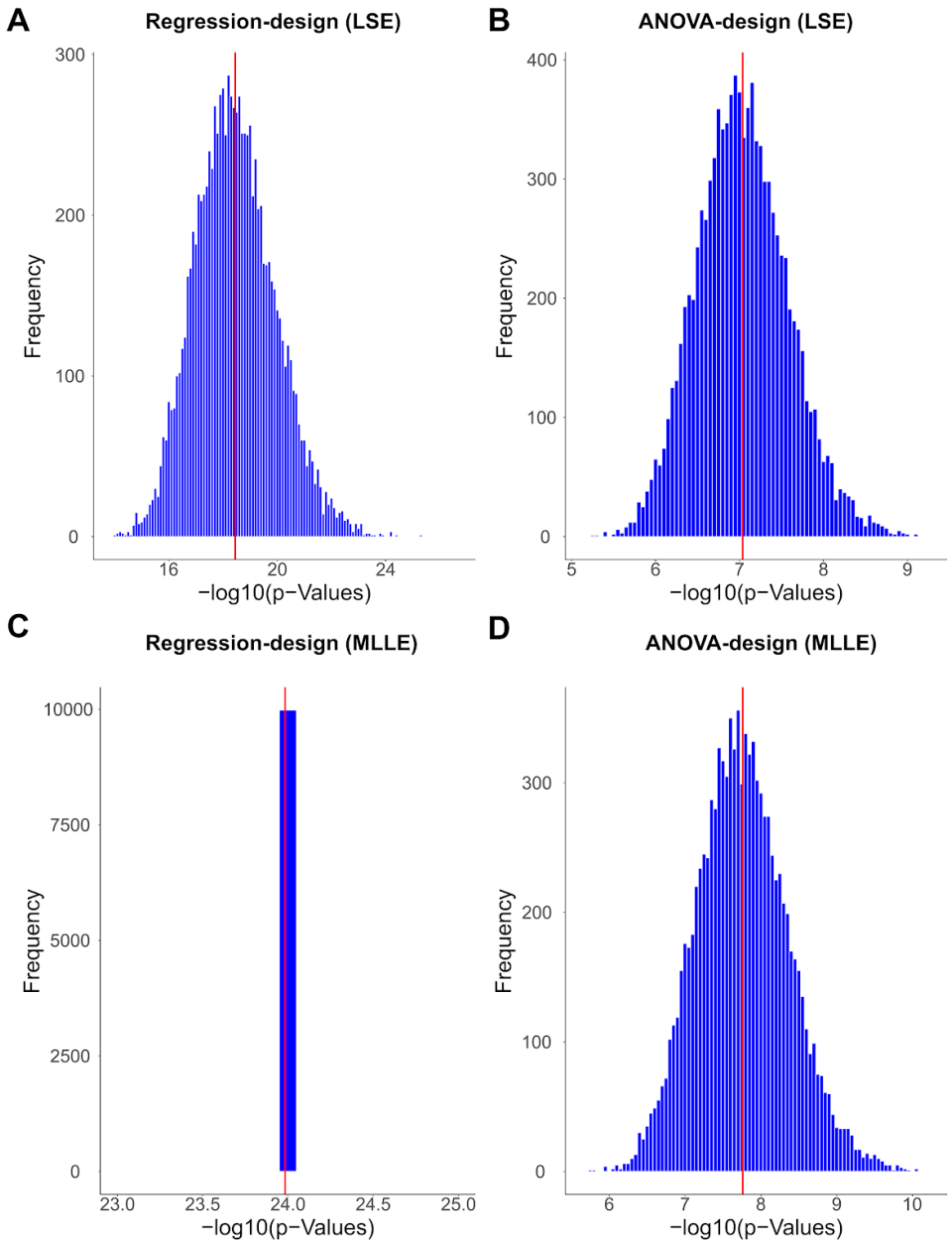

Figure 4: Negative log10 p-value distributions for regression-design and ANOVA-design for both MLLE and LSE methods for two-level and $\sigma = 1$ simulations. Across two-level simulations for $\sigma = 1$, the negative log, base ten p-value distributions were plotted for regression-design and ANOVA-design for MLLE and LSE methods to compare statistical power and to observe any potential bias. Average negative log base ten p-values across the simulations are indicated by the red line. A) Regression-design p-value distribution for LSE methods with a mean p-value of 2.738349e-17 ± 5.3676e-16. B) ANOVA-design p-value distribution for LSE methods with a mean p-value of 1.885066e-07 ± 5.726072e-07. C) Regression-design p-value distribution for MLLE methods with a mean p-value of 3.552714e-19 ± 1.793144e-17. Because the calculated p-values for these simulations were below machine epsilon (and thus the logarithmic value could not be calculated for them), a value of 1e-24 was added to each value to plot the distribution. This was not done when calculating the average p-value nor the 95% confidence interval. D) ANOVA-design p-value distributions for MLLE methods with a mean p-value of 4.160012e-08 ± 1.492111e-07. (± values are calculated as 95% confidence intervals)

Two Level:

	σ = 1		σ = 2		σ = 4		σ = 6		σ = 8		σ = 12		σ = 16		σ = 24	
LSE	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA
$\bar{P}$ $\bar{P}$	2.74E-17	1.89E-07	7.11E-11	3.46E-06	1.26E-05	0.00053	0.001495	0.00924	0.014749	0.039599	0.093067	0.144578	0.190955	0.245854	0.322068	0.360554
$(P<0.05)_{count}$	10000	10000	10000	10000	10000	9999	9962	9599	9279	8086	6242	4843	3999	3083	2055	1639

	σ = 1		σ = 2		σ = 4		σ = 6		σ = 8		σ = 12		σ = 16		σ = 24	
MLLE	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA
$\bar{P}$ $\bar{P}$	3.55E-19	4.16E-08	1.00E-11	1.14E-06	5.72E-06	0.00031	0.001019	0.007136	0.011925	0.033626	0.08361	0.132196	0.17715	0.229746	0.305887	0.337842
$(P<0.05)_{count}$	10000	10000	10000	10000	10000	9999	9976	9711	9434	8433	6719	5360	4425	3548	2390	1993

Four Level:

	σ = 1		σ = 2		σ = 4		σ = 6		σ = 8		σ = 12		σ = 16		σ = 24	
LSE	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA
$\bar{P}$	2.93E-17	2.96E-10	6.39E-11	1.64E-07	1.28E-05	0.000398	0.001517	0.011051	0.014311	0.053217	0.092372	0.176324	0.190337	0.282396	0.321119	0.38242
$(P<0.05)_{count}$	10000	10000	10000	10000	10000	9998	9958	9484	9334	7433	6372	3868	4043	2270	2124	1264

	σ = 1		σ = 2		σ = 4		σ = 6		σ = 8		σ = 12		σ = 16		σ = 24	
MLLE	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA
$\bar{P}$	3.77E-19	6.94E-12	8.49E-12	1.42E-08	5.78E-06	0.000139	0.00104	0.006281	0.011527	0.037323	0.082879	0.141429	0.17685	0.237826	0.305086	0.330299
$(P<0.05)_{count}$	10000	10000	10000	10000	10000	9999	9971	9727	9461	8228	6824	4957	4474	3157	2447	1953

Eight Level:

	σ = 1		σ = 2		σ = 4		σ = 6		σ = 8		σ = 12		σ = 16		σ = 24	
LSE	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA
$\bar{P}$	2.80E-17	8.77E-10	5.04E-11	1.85E-06	1.07E-05	0.002341	0.001425	0.033265	0.014978	0.107075	0.092951	0.249264	0.188388	0.337825	0.31722	0.422138
$(P<0.05)_{count}$	10000	10000	10000	10000	10000	9954	9960	8174	9260	5303	6319	2383	4105	1525	2137	908

	σ = 1		σ = 2		σ = 4		σ = 6		σ = 8		σ = 12		σ = 16		σ = 24	
MLLE	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA	Lin. Reg.	ANOVA
$\bar{P}$	3.55E-19	2.96E-13	6.09E-12	1.82E-08	4.71E-06	0.000318	0.000966	0.010597	0.012117	0.050269	0.083501	0.146813	0.174781	0.219479	0.301311	0.290823
$(P<0.05)_{count}$	10000	10000	10000	10000	10000	9994	9974	9482	9425	7719	6752	4650	4521	3397	2442	2377

Figure 5. Average p-values and total significant p-value counts for all simulations with all combinations of levels, sigmas, design methods, and parameter estimation methods. For each level that was simulated, the average p-value and total significant ($p < 0.05$) p-values were tabulated and compared regression-design (Lin. Reg.) to ANOVA-design (ANOVA) for both LSE and MLLE parameter estimation methods. Comparison across a row demonstrates that as σ increases, the average p-value increases and the total number of significant p-values decreases. Comparison down a column, demonstrates that regardless of level, similar average p-values and total significant p-value counts are observed. Comparison of adjacent Lin. Reg. and ANOVA columns for any given level and σ demonstrate that regardless of experiment design, both methods are viable at detecting a treatment effect. From this we can conclude that regardless of level (two, four or eight), design method (ANOVA or regression), or parameter estimation method (LSE or MLLE), there is equal statistical power. The only affecting factor of statistical power is σ. Average p-value gradients: green color indicates a significant average p-value with increasing gradient towards yellow until $p = 0.05$. For $p > 0.05$, the gradient continues increasingly towards red. Total p-count gradients: green color indicates that all simulations demonstrated significance with an increasing gradient towards yellow to red and the total number of significant count decreased.

References

[1] Hryhorowicz, M., Zeyland, J., Słomski, R., & Lipiński, D. (2017). Genetically Modified Pigs as Organ Donors for Xenotransplantation. Molecular Biotechnology, 59(9), 435–444. https://doi.org/10.1007/s12033-017-0024-9

[2] Factors affecting the gestation period of pigs in Nigeria. - PubMed - NCBI. (n.d.). Retrieved December 3, 2018, from https://www.ncbi.nlm.nih.gov/pubmed/7233562

[3] Jones, Stuart. (2018). Retrieved from https://github.com/fjhuizar/biocomputing_StatsGroupProject

[4] 4.4.3.1. Least Squares. (n.d.). Retrieved December 3, 2018, from https://www.itl.nist.gov/div898/handbook/pmd/section4/pmd431.htm

[5] 8.4.1.2. Maximum likelihood estimation. (n.d.). Retrieved December 3, 2018, from https://www.itl.nist.gov/div898/handbook/apr/section4/apr412.htm

[6] Wan, A. (n.d.). Assumptions and Properties of Ordinary Least Squares, and Inference in the Linear Regression Model. Retrieved December 3, 2018, from http://personal.cb.cityu.edu.hk/msawan/teaching/FB8916/FB8916Ch2.pdf

[7] Gray-Steinhauer, L. (n.d.). ANOVA Assumptions. Retrieved December 3, 2018, from https://sites.ualberta.ca/~lkgray/uploads/7/3/6/2/7362679/slides_-_anova_assumptions.pdf

[8] F-test for Regression. (n.d.). Retrieved December 3, 2018, from http://facweb.cs.depaul.edu/sjost/csc423/documents/f-test-reg.htm

[9] Stats: One-Way ANOVA. (n.d.). Retrieved December 3, 2018, from https://people.richland.edu/james/lecture/m170/ch13-1wy.html

[10] Likelihood Ratio: Wilks's Theorem. (n.d.). Retrieved December 3, 2018, from https://stephens999.github.io/fiveMinuteStats/wilks.html

[11] Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. The Annals of Mathematical Statistics, 9(1), 60–62. https://doi.org/10.1214/aoms/1177732360