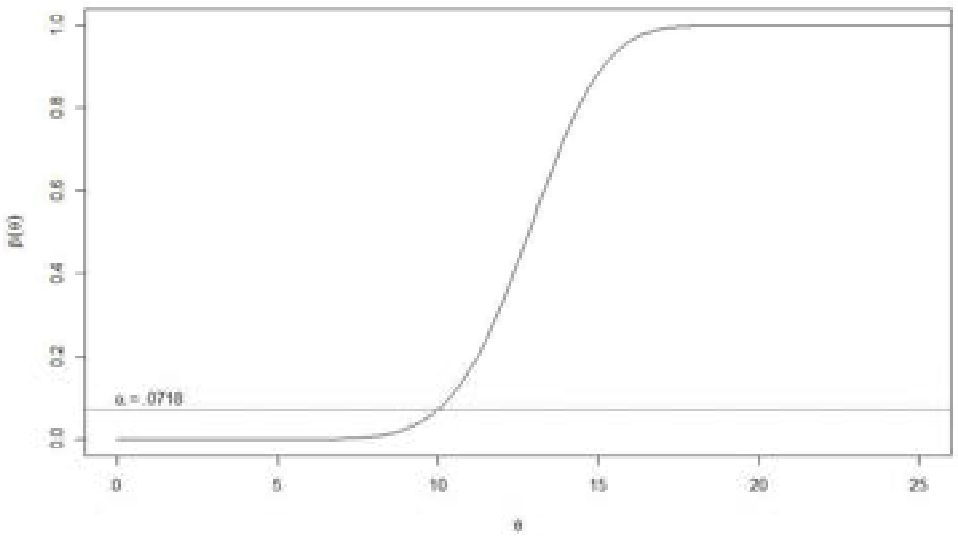
1. An ecologist suspects the mean rate of diseased pine trees per one acre, θ , is underestimated. To check this claim, the ecologist randomly selects 23 different areas: 5 areas each of size 0.25 acres, 10 areas each of size 1 acre, and 8 each of size 0.75 acres. The number of diseased pine trees in each randomly selected area is recorded and assumed to be independent of one another. To test the proposed claim, the ecologist is presented with the following hypothesis test function:

$$\phi_M(\boldsymbol{X}) = \begin{cases} 1 & X_{(n)} > c \\ 0 & \text{else} \end{cases}$$

The following plot of the power function for ϕ_M was produced, and a horizontal line was drawn at the significance level, $\alpha = 0.07183788$.



(a) Explain why a Poisson($a_i\theta$) distribution would be a reasonable model for the random variables $X_1, \ldots, X_n$. Your explanation should also include **in words** what n, $X_1, \ldots, X_n$, $a_1, \ldots, a_n$,θ, and $x_1, \ldots, x_n$ are in this situation, and, if applicable, give the numerical values of each.

(b) Based on the power function provided in the plot, what are the hypotheses the ecologist wants to test? Be sure to include any appropriate numerical values and use words to describe any symbols used.

(c) Derive the power function for ϕ_M that is plotted. Show and explain all work.

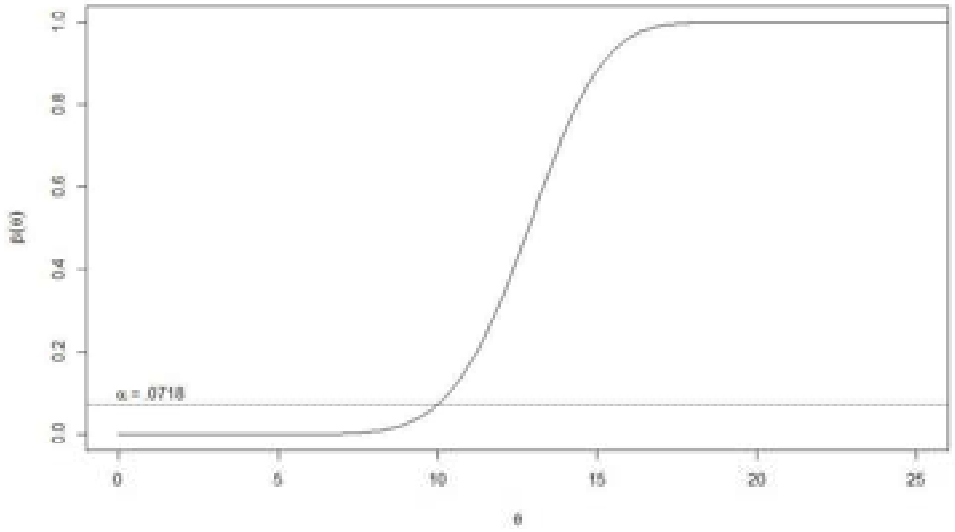
(d) Explain in words how you would find c.

(e) Explain why $\alpha = 0.07183788$ instead of a more typical value such as 0.05 or 0.10.

(f) The ecologist is presented with another hypothesis test function:

$$\phi_S(\boldsymbol{X}) = \begin{cases} 1 & \sum_{i=1}^{n} X_i > 191 \\ 0 & \text{else.} \end{cases}$$

- **On the plot on provided on page 3**, sketch what you think the power function of this hypothesis test would look like in relation to the power function of ϕ_M. Assume approximately the same significance level is used. (Note: If you are unable to sketch the power function on a printed copy of the plot, please trace the plot on a blank sheet of paper and clearly label each power function.)
- **In words**, explain thoroughly why the power function for ϕ_S should look similar to what you sketched. A complete explanation should also include rationale for the relative positions chosen for the power functions over the entire parameter space.


β(θ)
0.0
0.2
0.4
0.6
0.8
1.0
α = .0718
0
5
10
15
20
25
θ

2. Suppose the annual income, X, of a randomly selected household follows a Pareto(θ) distribution with pdf

$$f_X(x) = \frac{\theta}{x^2}; \quad x > \theta.$$

Let $X_1, \ldots, X_n$ be a random sample of households. Economists would like to make inferences about the parameter θ.

(a) Explain what θ means in this context.

(b) Derive the likelihood ratio test (LRT) for testing $H_0 : \theta = \theta_0$ versus $H_0 : \theta \neq \theta_0$ at the .01 significance level. Show all work and specify an exact critical value for the test. Show all work.

(c) Suppose the following household incomes (in dollars) were observed for 15 randomly selected households:

Household Income				
78,471	63,088	42,323	54,043	37,854
70,340	32,766	29,258	242,645	24,666
39,217	21,961	37,763	49,781	44,250

Use these data and the test you derived in part b) to test the hypotheses of interest at the 0.01 significance level and with a null value of \$20,000. Be sure to show and explain all work, and state your decision in a meaningful way that uses the given context.

(d) The economists are also interested in obtaining a $(1-\alpha)100\%$ confidence interval for θ .

- Derive a two-sided $(1-\alpha)100\%$ confidence interval for θ that is based on a sufficient statistic. Show and explain all work.
- Use the data given in part c) and your derivations to obtain a two-sided 99% confidence interval for θ . Interpret the interval in the context of the scenario. What does the interval suggest about the hypotheses of interest?

3. Let $\boldsymbol{Y}_{ij}$ be independent multinomial$(1, p_{i1}, \ldots, p_{iC})$ random variables where $i = 1, \ldots, T$ and $j = 1, \ldots, n$. That is $\boldsymbol{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijC})$ is a vector of length C with one element equal to 1 and the remaining elements are equal to 0.

(a) Derive MLE estimates of p_{ik}.

(b) Further, suppose that the vectors $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{iC})$ are independent Dirichlet random variables with pdf

$$\pi(\boldsymbol{p}_i) = \frac{\Gamma(\sum_{k=1}^{C} \alpha_k)}{\prod_{k=1}^{C} \Gamma(\alpha_k)} \prod_{k=1}^{C} p_{ik}^{\alpha_k - 1}; \quad 0 < p_{ik} < 1 \text{ and } \sum_{k=1}^{C} p_{ik} = 1.$$

Derive the joint posterior distribution of $\boldsymbol{p}_1, \ldots \boldsymbol{p}_T$.

4. You have had an initial meeting with a graduate student in Plant Pathology. Below are your notes from the initial client meeting along with the results for the analysis you ran in SAS. You know your client has taken 802 and they say they are familiar with SAS output and asked to see just the SAS output before the first follow up meeting. After you sent them the results, they email you back specific questions that they would like to discuss during the follow-up. Below are the following:

    1. The initial notes you took
    2. The SAS output that you sent the researcher
    3. The list of questions the graduate student emailed you that they would like to go over during the follow up meeting.

    Go through each question and write up a summary of how you would try to answer these questions during your follow-up meeting with the graduate student.

    Initial Meeting Notes:

    The experiment was conducted at four fields with center-pivot irrigation. Center pivot irrigators apply water in a circle. The units can be adjusted so that different amounts of irrigation can be applied in concentric rings.

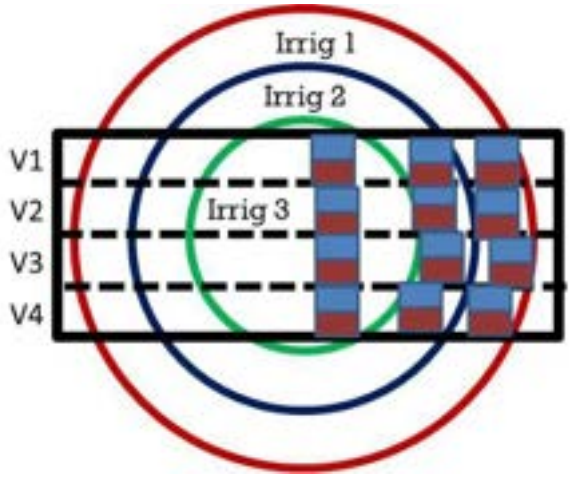
    Three irrigation levels: 1, 2, 3

    Four plant varieties: 1, 2, 3, 4 - planted in strips across the field.

    Disease is measured by "percent leaf area affected," where 0 means the leaf shows no disease symptoms and 1 means the leaf is completely damaged by the disease – no healthy leaf tissue remains.

    Research question: Is there a difference in the four varieties in "resistance" and does irrigation level impact the difference in resistance.

Figure 1: Schematic diagram of the design layout for an example field. Varieties and irrigation levels were randomized at any given field. The two-colored rectangles within each irrigation × variety zone are plots. Within each plot, several plants are sampled (not the same number in every plot). "Percent leaf area affected" is recorded on a per plant basis.



Results

Analysis (percent leaf area affected)

Model Information	
Data Set	WORK.PLANT
Response Variable	leaf_area
Response Distribution	Beta
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	field
Estimation Technique	Maximum Likelihood
Likelihood Approximation	Laplace
Degrees of Freedom Method	Containment

Class Level Information		
Class	**Levels**	**Values**
field	4	1 2 3 4
irrig	3	1 2 3
variety	4	1 2 3 4

Number of Observations Read	219
Number of Observations Used	219

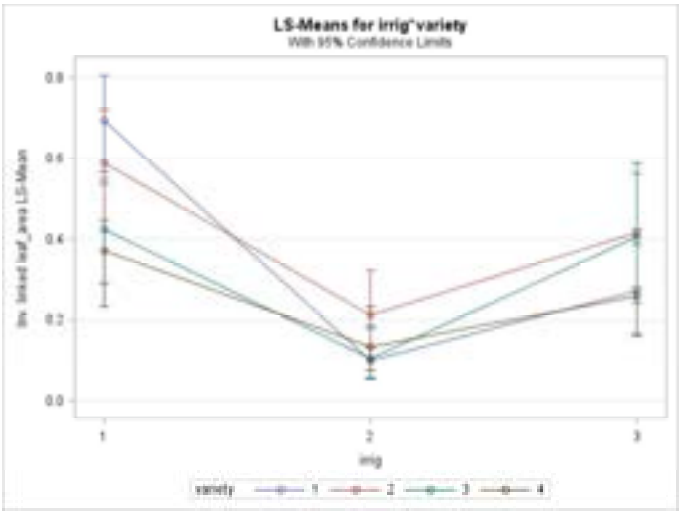
Fit Statistics for Conditional Distribution	
-2 log L(leaf_area | r. effects)	-287.16
Pearson Chi-Square	193.95
Pearson Chi-Square / DF	0.89

Covariance Parameter Estimates			
Cov Parm	**Subject**	**Estimate**	**Standard Error**
Intercept	**field**	0.09920	0.09841
irrig	**field**	0.02267	0.03572
variety	**field**	0.06350	0.05323
irrig*variety	**field**	0.000783	0.04817
Scale		5.7014	0.6113

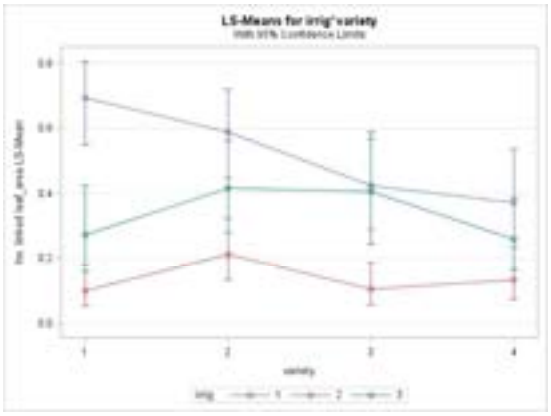
Solutions for Fixed Effects							
Effect	**irrig**	**variety**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept			-1.0427	0.2770	3	-3.76	0.0328
irrig	**1**		0.5209	0.3089	6	1.69	0.1428
irrig	**2**		-0.8146	0.3090	6	-2.64	0.0387
irrig	**3**		0	.	.	.	.
variety		**1**	0.05846	0.3460	9	0.17	0.8696
variety		**2**	0.7009	0.3105	9	2.26	0.0504
variety		**3**	0.6580	0.3772	9	1.74	0.1151
variety		**4**	0	.	.	.	.
irrig*variety	**1**	**1**	1.2750	0.4300	18	2.97	0.0083
irrig*variety	**1**	**2**	0.1841	0.3844	18	0.48	0.6378
irrig*variety	**1**	**3**	-0.4427	0.4391	18	-1.01	0.3267
irrig*variety	**1**	**4**	0	.	.	.	.
irrig*variety	**2**	**1**	-0.4004	0.4434	18	-0.90	0.3785
irrig*variety	**2**	**2**	-0.1518	0.3775	18	-0.40	0.6924
irrig*variety	**2**	**3**	-0.9521	0.4672	18	-2.04	0.0565
irrig*variety	**2**	**4**	0	.	.	.	.
irrig*variety	**3**	**1**	0	.	.	.	.
irrig*variety	**3**	**2**	0	.	.	.	.
irrig*variety	**3**	**3**	0	.	.	.	.
irrig*variety	**3**	**4**	0	.	.	.	.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
irrig	2	6	55.51	0.0001
variety	3	9	3.24	0.0746
irrig*variety	6	18	3.99	0.0103

irrig*variety Least Squares Means													
irrig	**variety**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**	**Mean**	**Standard Error Mean**	**Lower Mean**	**Upper Mean**
1	**1**	0.8117	0.2926	18	2.77	0.0125	0.05	0.1969	1.4264	0.6925	0.06231	0.5491	0.8063
1	**2**	0.3632	0.2744	18	1.32	0.2022	0.05	-0.2133	0.9397	0.5898	0.06639	0.4469	0.7190
1	**3**	-0.3066	0.2766	18	-1.11	0.2823	0.05	-0.8877	0.2746	0.4239	0.06755	0.2916	0.5682
1	**4**	-0.5218	0.3180	18	-1.64	0.1182	0.05	-1.1900	0.1463	0.3724	0.07433	0.2333	0.5365
2	**1**	-2.1993	0.3240	18	-6.79	<.0001	0.05	-2.8799	-1.5187	0.09982	0.02911	0.05316	0.1797
2	**2**	-1.3082	0.2675	18	-4.89	0.0001	0.05	-1.8702	-0.7462	0.2128	0.04481	0.1335	0.3216
2	**3**	-2.1515	0.3212	18	-6.70	<.0001	0.05	-2.8263	-1.4766	0.1042	0.02998	0.05592	0.1859
2	**4**	-1.8573	0.3191	18	-5.82	<.0001	0.05	-2.5278	-1.1869	0.1350	0.03727	0.07393	0.2338
3	**1**	-0.9843	0.3239	18	-3.04	0.0071	0.05	-1.6648	-0.3037	0.2720	0.06415	0.1591	0.4247
3	**2**	-0.3418	0.2852	18	-1.20	0.2463	0.05	-0.9410	0.2574	0.4154	0.06926	0.2807	0.5640
3	**3**	-0.3848	0.3565	18	-1.08	0.2947	0.05	-1.1337	0.3641	0.4050	0.08590	0.2435	0.5900
3	**4**	-1.0427	0.2770	18	-3.76	0.0014	0.05	-1.6247	-0.4608	0.2606	0.05338	0.1646	0.3868



Simple Effect Comparisons of irrig*variety Least Squares Means By irrig Adjustment for Multiple Comparisons: Tukey-Kramer																		
Simple Effect Level	**variety**	**_variety**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Adj P**	**Alpha**	**Lower**	**Upper**	**Adj Lower**	**Adj Upper**	**Odds Ratio**	**Lower Odds Ratio**	**Upper Odds Ratio**	**Adj Lower Odds Ratio**	**Adj Upper Odds Ratio**
irrig 1	**1**	**2**	0.4485	0.3156	18	1.42	0.1724	0.5030	0.05	-0.2145	1.1115	-0.4434	1.3404	1.566	0.807	3.039	0.642	3.821
irrig 1	**1**	**3**	1.1183	0.3188	18	3.51	0.0025	0.0122	0.05	0.4485	1.7881	0.2172	2.0193	3.060	1.566	5.978	1.243	7.533
irrig 1	**1**	**4**	1.3335	0.3558	18	3.75	0.0015	0.0073	0.05	0.5860	2.0810	0.3279	2.3391	3.794	1.797	8.012	1.388	10.371
irrig 1	**2**	**3**	0.6698	0.3015	18	2.22	0.0394	0.1552	0.05	0.03631	1.3032	-0.1824	1.5219	1.954	1.037	3.681	0.833	4.581
irrig 1	**2**	**4**	0.8850	0.3403	18	2.60	0.0181	0.0776	0.05	0.1701	1.5999	-0.07672	1.8467	2.423	1.185	4.953	0.926	6.339
irrig 1	**3**	**4**	0.2152	0.3389	18	0.64	0.5333	0.9193	0.05	-0.4967	0.9272	-0.7426	1.1730	1.240	0.609	2.527	0.476	3.232
irrig 2	**1**	**2**	-0.8911	0.3346	18	-2.66	0.0159	0.0689	0.05	-1.5941	-0.1880	-1.8368	0.05468	0.410	0.203	0.829	0.159	1.056
irrig 2	**1**	**3**	-0.04780	0.3741	18	-0.13	0.8997	0.9992	0.05	-0.8337	0.7381	-1.1051	1.0095	0.953	0.434	2.092	0.331	2.744
irrig 2	**1**	**4**	-0.3419	0.3767	18	-0.91	0.3760	0.8010	0.05	-1.1334	0.4495	-1.4066	0.7228	0.710	0.322	1.568	0.245	2.060
irrig 2	**2**	**3**	0.8433	0.3325	18	2.54	0.0207	0.0877	0.05	0.1447	1.5419	-0.09654	1.7831	2.324	1.156	4.673	0.908	5.948
irrig 2	**2**	**4**	0.5491	0.3306	18	1.66	0.1140	0.3717	0.05	-0.1453	1.2436	-0.3851	1.4834	1.732	0.865	3.468	0.680	4.408
irrig 2	**3**	**4**	-0.2941	0.3722	18	-0.79	0.4397	0.8580	0.05	-1.0762	0.4879	-1.3462	0.7579	0.745	0.341	1.629	0.260	2.134
irrig 3	**1**	**2**	-0.6424	0.3532	18	-1.82	0.0856	0.2970	0.05	-1.3844	0.09953	-1.6406	0.3557	0.526	0.250	1.105	0.194	1.427
irrig 3	**1**	**3**	-0.5995	0.4129	18	-1.45	0.1637	0.4851	0.05	-1.4669	0.2679	-1.7663	0.5674	0.549	0.231	1.307	0.171	1.764
irrig 3	**1**	**4**	0.05846	0.3460	18	0.17	0.8677	0.9982	0.05	-0.6684	0.7853	-0.9193	1.0362	1.060	0.513	2.193	0.399	2.819
irrig 3	**2**	**3**	0.04295	0.3837	18	0.11	0.9121	0.9995	0.05	-0.7631	0.8490	-1.0414	1.1274	1.044	0.466	2.337	0.353	3.087
irrig 3	**2**	**4**	0.7009	0.3105	18	2.26	0.0367	0.1458	0.05	0.04847	1.3533	-0.1768	1.5786	2.016	1.050	3.870	0.838	4.848
irrig 3	**3**	**4**	0.6580	0.3772	18	1.74	0.0982	0.3312	0.05	-0.1346	1.4505	-0.4082	1.7241	1.931	0.874	4.265	0.665	5.608



Simple Effect Comparisons of irrig*variety Least Squares Means By variety Adjustment for Multiple Comparisons: Tukey-Kramer																		
Simple Effect Level	**irrig**	**_irrig**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Adj P**	**Alpha**	**Lower**	**Upper**	**Adj Lower**	**Adj Upper**	**Odds Ratio**	**Lower Odds Ratio**	**Upper Odds Ratio**	**Adj Lower Odds Ratio**	**Adj Upper Odds Ratio**
variety 1	**1**	**2**	3.0109	0.3402	18	8.85	<.0001	<.0001	0.05	2.2963	3.7256	2.1428	3.8791	20.307	9.937	41.496	8.523	48.381
variety 1	**1**	**3**	1.7959	0.3340	18	5.38	<.0001	0.0001	0.05	1.0943	2.4976	0.9436	2.6483	6.025	2.987	12.153	2.569	14.130
variety 1	**2**	**3**	-1.2150	0.3516	18	-3.46	0.0028	0.0076	0.05	-1.9537	-0.4763	-2.1124	-0.3176	0.297	0.142	0.621	0.121	0.728
variety 2	**1**	**2**	1.6714	0.2580	18	6.48	<.0001	<.0001	0.05	1.1293	2.2134	1.0129	2.3298	5.320	3.094	9.147	2.754	10.276
variety 2	**1**	**3**	0.7050	0.2748	18	2.57	0.0195	0.0487	0.05	0.1277	1.2823	0.003662	1.4063	2.024	1.136	3.605	1.004	4.081
variety 2	**2**	**3**	-0.9664	0.2673	18	-3.62	0.0020	0.0053	0.05	-1.5279	-0.4048	-1.6486	-0.2842	0.380	0.217	0.667	0.192	0.753
variety 3	**1**	**2**	1.8449	0.3133	18	5.89	<.0001	<.0001	0.05	1.1867	2.5031	1.0453	2.6444	6.327	3.276	12.220	2.844	14.076
variety 3	**1**	**3**	0.07818	0.3494	18	0.22	0.8255	0.9728	0.05	-0.6560	0.8123	-0.8136	0.9700	1.081	0.519	2.253	0.443	2.638
variety 3	**2**	**3**	-1.7667	0.3840	18	-4.60	0.0002	0.0006	0.05	-2.5735	-0.9600	-2.7468	-0.7867	0.171	0.076	0.383	0.064	0.455
variety 4	**1**	**2**	1.3355	0.3482	18	3.84	0.0012	0.0033	0.05	0.6040	2.0671	0.4468	2.2242	3.802	1.829	7.902	1.563	9.246
variety 4	**1**	**3**	0.5209	0.3089	18	1.69	0.1090	0.2375	0.05	-0.1281	1.1699	-0.2675	1.3093	1.684	0.880	3.222	0.765	3.704
variety 4	**2**	**3**	-0.8146	0.3090	18	-2.64	0.0168	0.0423	0.05	-1.4638	-0.1654	-1.6033	-0.02601	0.443	0.231	0.848	0.201	0.974

Researcher Questions

These are the questions that are going to guide your discussion within the follow-up meeting. Write short summaries of how you would answer the graduate student's questions and what you would be prepared to discuss in your meeting.

(a) There were 219 observations. Why are there so few degrees of freedom for error for testing the interaction term?

(b) What is the scale parameter in the covariance parameter table?

(c) Why is the response function "Beta" and link function "Logit"?

(d) Why are some of the solutions for the fixed effects zero and have no standard errors or p-values?

(e) I know that the denominator degrees of freedom different for the different factors in the "Type III" table are different because of different error terms. How do I determine what error term is used for each effect?

(f) In the Least Squares means tables, how can we have negative "Estimate" values and "Estimate" values that are greater than 1?

(g) In the Least Squares means table, how do we interpret the "Mean" column?

(h) How do I interpret the p-values in the Least Squares means table?

(i) What does "simple effect" mean in the Least Squares means by irrigation and Least Squares means by variety tables?

(j) What is an adjusted p-value? Why would we want to use this?

(k) What are odds ratios and how do we interpret them?

(l) Can I look at variety and irrigation level individually?

- Why or why not? In the second figure, the lines don't cross... I heard from someone in my department that I can look at the individual variables when the lines don't cross.

(m) How do I interpret the bars in the plots?

1. There is a data set from a multi-location study, with 2 treatments and 10 locations. Treatment 0 is a standard treatment and Treatment 1 is an experimental treatment whose purpose is to reduce the number of defective connections in a manufactured item. Ideally, this number should be zero, but in practice, a few defective connections are inevitable. The product is designed to work around them, but fewer defects translate to greater accuracy and improved efficiency. In the data set the response variable is denoted by `COUNT`. The data set can be found in data.sas. (Note that when doing this kind of evaluation, the company is required by law to provide broad inference space estimates that represent all of it production facilities (`LOCATIONS`) worldwide.)

   (a) Analyze the data in 4 different ways:

   - standard ANOVA on `COUNT`
   - ANOVA on the log transform log(`COUNT`+1)
   - Generalized linear mixed model assuming Poisson distribution
   - Generalized linear mixed model assuming negative binomial distribution

   For each analysis,

   - Write out the model. Define each terms and state assumptions. record results for the following
   - test of $H_o : \tau_0 = \tau_1$
   - point and interval estimates of λ_0 and λ_1 , the data scale treatment means

   (b) **Write a short report** summarizing the key results from the analysis using relevant SAS or R output. Make sure you compare the analyses and results and you discuss which one you would consider and why.

   Make sure you attach your SAS/R code in the Appendix in such a way that I can run your code without modifying anything.

2. Suppose $Y_1, \ldots, Y_n$ are independent normal random variables with mean μ and variance σ^2. One can show that $X = (n-1)S^2/\sigma^2$ has a χ^2 distribution with n–1 degrees of freedom, where S^2 is the sample variance.

   Complete the following problems. Use dynamic document creation via the knitr package to help make your results reproducible. Turn in your corresponding source file (e.g., LaTeX, LyX, or R Markdown) along with a PDF produced from the file.

   (a) Why is X a pivotal quantity?

   (b) A commonly used $(1-\alpha)100\%$ confidence interval for σ^2 is

   $$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}$$

   where $\chi^2_{1-\alpha/2,n-1}$ is the $1-\alpha/2$ quantile from a χ^2 distribution with n–1 degrees of freedom. Derive this interval with the help of X being a pivotal quantity.

   (c) Suppose $\mu = 0$ and $\sigma^2 = 1$. Compute the estimated true confidence level (i.e., coverage level) for the interval in 2b. Use a 95% confidence level, $n = 100$, and $R = 1,000$ simulated data sets for your computations. Make sure to set a seed number so that your exact same samples can be recovered.

   (d) Based on your results from 2c, is the confidence interval performing as expected?

3. Suppose $Y_1, \ldots, Y_n$ are now independent logistic random variables with mean equal to 0 and variance equal to 1,

   Note: The pdf is

   $$f(x) = \frac{1}{\beta}\frac{e^{-x/\beta}}{[1+e^{-x/\beta}]^2}; \quad -\infty < x < \infty$$

   with $\beta = \sqrt{3}/\pi$.

   (a) Compute the estimated true confidence level (i.e., coverage level) for the interval in 2b. Use a 95% confidence level, $n = 100$, and $R = 1,000$ simulated data sets for your computations. Make sure to set a seed number so that your exact same samples can be recovered.

   (b) Repeat the computations in 3a for other sample sizes lower and higher than n = 100. Describe trends that are present.

   (c) You should see from 3a and 3b that the confidence interval is not performing well. Why does this occur?

   (d) Should this confidence interval be recommended for general practice? Explain.