ST512

Fall Semester, 2008

Quiz 2

Name:____

Directions: Answer questions as directed. Please show work. Partial credit may be awarded for correct expressions given in incomplete answers. To save time, it is ok to give answers as numeric expressions without carrying out every last operation. If the question asked for the standard error of the estimated mean of a population at $x = 14$ in a simple linear regression, the following response would receive full credit:

$$\widehat{SE}(\hat{\beta}_0 + 14\hat{\beta}_1) = \sqrt{16.4\left(\frac{1}{38} + \frac{(14 - 10.8)^2}{190}\right)}$$

You may assume a level of significance of $\alpha = .05$ and a confidence level of 95% for any tests or confidence intervals.

There is a blank page at the end for extra space. You may also use the back of the page if necessary.

1. Consider a balanced experiment with a completely randomized design to test for the effect of a factor with $t = 4$ levels. Assume the usual model for the response on experimental unit j ($j = 1, 2, \ldots, n$) randomized to treatment i:

$$Y_{ij} = \mu + \tau_i + E_{ij}$$

where E_{ij} are assumed i.i.d. $N(0, \sigma^2)$.

Circle T for true or F for false, or provide a short answer:

(a) T or (F): the expectation of the error mean square is $E[MS(E)] = \sigma^2 + n\sum\tau_i^2$

(b) (T) or F: As n increases, the type II error rate goes down.

(c) (T) or F: As σ increases, the type II error rate goes up.

(d) T or (F): For simultaneous 95% confidence intervals for the six differences among the four treatment means, the Bonferroni procedure will yield more narrow confidence intervals than the Tukey procedure.

(e) Suppose that interest lies in tests of $k = 5$ contrasts involving the four treatment means.

i. Give the definition of the familywise error rate in this context.
Pr(at least one type I error)

ii. Suppose $n = 5$, which procedure will lead to more narrow simultaneous 95% confidence intervals, Scheffe or Bonferroni? Since $t(.025/k) < \sqrt{(4-1)F(.05, 3, 16)}$, the intervals based on the Bonferroni procedure will be more narrow.

2. Twenty pigs were randomized to $t = 4$ diets to determine the effects of two protein sources on bodyweight. Daily weight gains (g) were measured with results below:

Diet	Sample size	Description	Sample mean	Sample variance	group
1	5	neither soybean nor corn gluten meal added	592	3320	
2	5	fixed amount, a, of soybean meal added	768	2520	
3	5	fixed amount, b, of corn gluten meal added	736	1280	
4	5	fixed amounts a of soybean meal and b of corn gluten meal are added.	724	10280	

The sum of squared deviations of all 20 observations about the grand mean is $SS[Tot] = 159900$. The sample variance among the four treatment means above is $1/(4-1)\sum_i(\bar{y}_{i+} - \bar{y}_{++})^2 = 6020$.

(a) Estimate the variance, σ^2, among daily weight gains for pigs on a fixed diet, with the usual assumption of homogeneity of variance. That is, report the $MS[E]$.

$$MS(E) = \frac{1}{4}(s_1^2 + s_2^2 + s_3^2 + s_4^2) = 4350$$

(b) Let the population mean weightgain among all animals consuming diet i be denoted μ_i. Conduct an F-test of the hypothesis that $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

$$F = \frac{MS(Trt)}{MS(E)} = \frac{6020 \times 5}{4350} = 6.9$$

$$F(.05, 3, 16) = 3.24$$

So H_0 may be rejected at level $\alpha = .05$. The sample treatment means differ significantly.

(c) Estimate the simple effect of adding soybean meal to the diet when corn gluten is not used ($\mu_2 - \mu_1$). Report an estimated standard error.

$$\bar{y}_{2+} - \bar{y}_{1+} = 176$$

$$\widehat{SE}(\bar{y}_{2+} - \bar{y}_{1+}) = \sqrt{MS(E)\frac{2}{5}} = 41.7$$

(d) Formulate and conduct a test of the hypothesis that the effect of adding soybean meal is the same regardless of whether corn gluten is also added to the diet.

$$H_0 : \mu_2 - \mu_1 = \mu_4 - \mu_3 (\text{i.e. no interaction})$$

$$\begin{aligned}
\hat{\theta} &= \bar{y}_{4+} - \bar{y}_{3+} - (\bar{y}_{2+} - \bar{y}_{1+}) \\
&= 188 \\
SS(\hat{\theta}) &= \frac{188^2}{\frac{4}{5}} \\
&= 44180 \\
F &= SS(\hat{\theta})/MS(E) \\
&= 10.2 \\
F(.05, 1, 16) &= 4.49
\end{aligned}$$

(e) Consider the simple effect of adding corn gluten meal to the diet when soybean meal is not used. Estimate the difference between this simple effect and the one estimated in part (c). Estimate the standard error for this difference.

$$\theta = \mu_2 - \mu_1 - (\mu_3 - \mu_1) = \mu_2 - \mu_3$$

$$\hat{\theta} = \bar{y}_2 - \bar{y}_3 = 32$$

$$\widehat{SE}(\hat{\theta}) = \sqrt{\frac{2}{5} MS(E)} = 41.7$$

(f) Obtain Tukey's honestly significant difference for all six pairwise comparisons among the four diet means. Add letters to the rightmost column in the table to indicate which pairs do not differ significantly, at familywise error rate .05.

$$HSD = q(.05, 4, 16)\sqrt{\frac{MS(E)}{5}} = 4.05\sqrt{\frac{4350}{5}} = 119$$

($\bar{y}_1$ differs significantly from the other three sample diet means.)

3. An experiment investigates the effect of seeding rates (90,110,130,150 and 170 $lb/acre$) on wheat yield. A field is divided into four rows of five plots each. Within each row, the five seeding rates are randomized to the five plots. In the summations given below, y_{ij} denotes the yield for treatment i assigned to row j:

$$\begin{aligned}
\sum_{i=1}^{5}\sum_{j=1}^{4}(y_{ij} - \bar{y}_{++})^2 &= 543.7 \\
\sum_{i=1}^{5}\sum_{j=1}^{4}(\bar{y}_{i+} - \bar{y}_{++})^2 &= 416.8 \\
\sum_{i=1}^{5}\sum_{j=1}^{4}(\bar{y}_{+j} - \bar{y}_{++})^2 &= 74.4 \\
\sum_{i=1}^{5}\sum_{j=1}^{4}(y_{ij} - \bar{y}_{i+} - \bar{y}_{+j} + \bar{y}_{++})^2 &= 52.6
\end{aligned}$$

with treatment means and standard deviations

Treatment (i)	Seeding rate	rows	mean ($\bar{y}_{i+}$)	std. dev.
1	90	4	47.7	3.9
2	110	4	53.1	3.9
3	130	4	55.8	1.9
4	150	4	60.8	0.6
5	170	4	58.8	2.7

(a) This is a randomized complete block design. What are the "blocks"? rows

(b) What is the block sum of squares? $SS[block]$ =? 74.4

(c) Estimate the mean difference in yield between levels 150 and 170 $lb/acre$. Report a standard error.

$$\begin{aligned}
\bar{y}_{5+} - \bar{y}_{4+} &= -2 \\
\widehat{SE}(\bar{y}_{5+} - \bar{y}_{4+}) &= \sqrt{MS(E)\frac{2}{4}} \\
&= \sqrt{\frac{52.6}{12}\frac{2}{4}} \\
&= \sqrt{4.83(2/4)} \\
&= 1.48
\end{aligned}$$

(d) Estimated linear and quadratic polynomial contrasts are given below.

$$\begin{aligned}\hat{\theta}_L &= -2\bar{y}_{1+} - \bar{y}_{2+} + 0\bar{y}_{3+} + \bar{y}_{4+} + 2\bar{y}_{5+} &= 29.7\\ \hat{\theta}_Q &= 2\bar{y}_{1+} - \bar{y}_{2+} + -2\bar{y}_{3+} - \bar{y}_{4+} + 2\bar{y}_{5+} &= -12.5\end{aligned}$$

i. Are these estimated contrasts orthogonal? What is their covariance, $\mathrm{Cov}(\hat{\theta}_L, \hat{\theta}_Q)$?
Yes they are orthogonal, and therefore have covariance 0.

ii. Test for lack of fit in a model in which mean yield depends linearly on seeding rate, (with different intercepts for different rows). Is the linear model ok?

$$\begin{aligned}SS(\hat{\theta}_L) &= \frac{\hat{\theta}_L^2}{10/4}\\ &= 352.8\\ F_{LOF} &= \frac{SS[LOF]/(5-1-1)}{MS(E)}\\ &= \frac{(416.8-352.8)/3}{52.6/12}\\ &= \frac{64/3}{4.38}\\ &= 4.86\\ F(.05,3,12) &= 3.49\end{aligned}$$

Since $F_{LOF} > F(.05,3,12)$, we reject the simple linear regression model in favor of a more complex polynomial.

(e) A reduced model is fit (SAS output below) that includes parameters for the effects of rows 1,2 and 3 ($\{\rho_j\}$), and only two parameters for a quadratic effect of seeding rate (x_i):

$$y_{ij} = \mu_4 + \rho_j + \beta_1 x_i + \beta_2 x_i^2 + E_{ij}.$$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	471.8648571	94.3729714	18.39	<.0001
Error	14	71.8371429	5.1312245		
Corrected Total	19	543.7020000			

Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		2.297500000 B	12.44551412	0.18	0.8562
row	1	-3.060000000 B	1.43265132	-2.14	0.0508
row	2	-1.120000000 B	1.43265132	-0.78	0.4474
row	3	-5.060000000 B	1.43265132	-3.53	0.0033
row	4	0.000000000 B	.	.	.
seedrate		0.728857143	0.19757013	3.69	0.0024
seedrate2		-0.002232143	0.00075676	-2.95	0.0106

i. For a given row, estimate the difference between seeding rate $x = 150$ and seeding rate $x = 170$ under the reduced model.

$$\hat{\beta}_1 * 150 + \hat{\beta}_2 * 150^2 - \hat{\beta}_1 * 170 + \hat{\beta}_2 * 170^2 = -0.3$$

ii. Test for lack of fit of the quadratic model.

$$\begin{aligned} F_{LOF} &= \frac{SS[LOF]/(5-2-1)}{MS(E)} \\ &= \frac{416.8 + 74.4 - 471.86)/2}{4.38} \\ &= \frac{19.3/2}{4.38} \\ &= 2.20 \\ F(.05, 2, 12) &= 3.89 \end{aligned}$$

Since F does not exceed the critical value, do not reject H_0. We may conclude that there is no significant lack of fit of the quadratic model.

The 95^{th} percentile for various F distributions

	Numerator degrees of freedom, $df1$									
$df2$	1	2	3	4	5	6	7	8	9	10
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165

The 95th percentile for studentized range (q) distributions
(The Tukey table)

	Numerator of treatment means, t									
Error *df*	2	3	4	5	6	7	8	9	10	11
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11
21	2.94	3.56	3.94	4.21	4.42	4.60	4.74	4.87	4.98	5.08
22	2.93	3.55	3.93	4.20	4.41	4.58	4.72	4.85	4.96	5.06
23	2.93	3.54	3.91	4.18	4.39	4.56	4.70	4.83	4.94	5.03
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01
25	2.91	3.52	3.89	4.15	4.36	4.53	4.67	4.79	4.90	4.99
26	2.91	3.51	3.88	4.14	4.35	4.51	4.65	4.77	4.88	4.98
27	2.90	3.51	3.87	4.13	4.33	4.50	4.64	4.76	4.86	4.96
28	2.90	3.50	3.86	4.12	4.32	4.49	4.62	4.74	4.85	4.94
29	2.89	3.49	3.85	4.11	4.31	4.47	4.61	4.73	4.84	4.93
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92

Critical values for Bonferroni t-tests to control the FWE at .05
(The Bonferroni table)

	Numerator of contrasts, k								
Error *df*	2	3	4	5	6	7	8	9	10
1	25.452	38.188	50.923	63.657	76.390	89.123	101.856	114.589	127.321
2	6.205	7.649	8.860	9.925	10.886	11.769	12.590	13.360	14.089
3	4.177	4.857	5.392	5.841	6.232	6.580	6.895	7.185	7.453
4	3.495	3.961	4.315	4.604	4.851	5.068	5.261	5.437	5.598
5	3.163	3.534	3.810	4.032	4.219	4.382	4.526	4.655	4.773
6	2.969	3.287	3.521	3.707	3.863	3.997	4.115	4.221	4.317
7	2.841	3.128	3.335	3.499	3.636	3.753	3.855	3.947	4.029
8	2.752	3.016	3.206	3.355	3.479	3.584	3.677	3.759	3.833
9	2.685	2.933	3.111	3.250	3.364	3.462	3.547	3.622	3.690
10	2.634	2.870	3.038	3.169	3.277	3.368	3.448	3.518	3.581
11	2.593	2.820	2.981	3.106	3.208	3.295	3.370	3.437	3.497
12	2.560	2.779	2.934	3.055	3.153	3.236	3.308	3.371	3.428
13	2.533	2.746	2.896	3.012	3.107	3.187	3.256	3.318	3.372
14	2.510	2.718	2.864	2.977	3.069	3.146	3.214	3.273	3.326
15	2.490	2.694	2.837	2.947	3.036	3.112	3.177	3.235	3.286
16	2.473	2.673	2.813	2.921	3.008	3.082	3.146	3.202	3.252
17	2.458	2.655	2.793	2.898	2.984	3.056	3.119	3.173	3.222
18	2.445	2.639	2.775	2.878	2.963	3.034	3.095	3.149	3.197
19	2.433	2.625	2.759	2.861	2.944	3.014	3.074	3.127	3.174
20	2.423	2.613	2.744	2.845	2.927	2.996	3.055	3.107	3.153
21	2.414	2.601	2.732	2.831	2.912	2.980	3.038	3.090	3.135
22	2.405	2.591	2.720	2.819	2.899	2.965	3.023	3.074	3.119
23	2.398	2.582	2.710	2.807	2.886	2.952	3.009	3.059	3.104
24	2.391	2.574	2.700	2.797	2.875	2.941	2.997	3.046	3.091
25	2.385	2.566	2.692	2.787	2.865	2.930	2.986	3.035	3.078
26	2.379	2.559	2.684	2.779	2.856	2.920	2.975	3.024	3.067
27	2.373	2.552	2.676	2.771	2.847	2.911	2.966	3.014	3.057
28	2.368	2.546	2.669	2.763	2.839	2.902	2.957	3.004	3.047
29	2.364	2.541	2.663	2.756	2.832	2.894	2.949	2.996	3.038
30	2.360	2.536	2.657	2.750	2.825	2.887	2.941	2.988	3.030