ID # 3104
BIOSTAT 536
HOMEWORK 3

1. We performed descriptive statistics on the overall sample and within groups based on sample quartiles of nadir PSA. For each baseline variable of interest, we calculated the total number of observations, the number of missing observations, the mean and standard deviation, and the minimum, median and maximum values when appropriate. The results are given in Table 1.

2. (a) We are interested in examining the association between relapse within 24 months and nadir PSA after adjusting for bone scan score and performance score. In Model 1, we fit a logistic regression model to the data, taking the indicator of relapse within 24 months as our response and nadir PSA as our predictor of interest. In addition, bone scan score and performance score were included as covariates in the model. According to this model, the estimated difference in log odds of relapse within 24 months between subjects whose nadir PSA differs by 1 ng/ml is 0.033, with greater nadir PSA being associated with greater log odds of relapse. A 95% confidence interval for this estimate is (-0.051, 0.116). Based on a P-value of 0.440, we lack sufficient evidence to reject the null hypothesis of no association between relapse within 24 months and nadir PSA.

   (b) In Model 2, we fit a logistic regression model to the data, taking the indicator of relapse within 24 months as our response and log nadir PSA as our predictor of interest. Bone scan score and performance score were also included as covariates in the model. After adjusting for bone scan score and performance score, the estimated difference in log odds of relapse within 24 months between subjects whose log nadir PSA differs by 1 unit is 0.844, with greater log nadir PSA being associated with greater odds of relapse. A 95% confidence interval for this estimate is (0.225, 1.462). Based on a P-value of 0.008, we have sufficient evidence to reject the null hypothesis of no association between log nadir PSA and relapse within 24 months.

   (c) In Model 3, we fit a logistic regression model to the data, taking the indicator of relapse within 24 months as our response and nadir PSA as our predictor of interest. Nadir PSA was modeled as a linear spline with knots at 1, 4, and 16 ng/ml. Bone scan score and performance score were also included as covariates in the model. To test whether there is evidence of an association between nadir PSA and relapse within 24 months, we performed a likelihood ratio test, comparing the full model described above to the model that only contains terms for bone scan score and performance score. Based on a P-value < 0.001, there is sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between nadir PSA and relapse within 24 months.

   (d) In Model 1, the intercept corresponds to the log odds of relapse within 24 months among subjects with a nadir PSA of 0 ng/ml, a bone scan score of 1, and a performance score of 0. In Model 2, the intercept corresponds to the log odds of relapse within 24 months among subjects with a nadir PSA of 1 ng/ml, a bone scan score of 1, and a performance score of

	Group	N	Missing	Mean	SD	Min	Median	Max
Age	Overall	50.00	0.00	67.44	5.77	58.00	66.00	86.00
	0-25% Percentile	15.00	0.00	67.27	5.46	58.00	66.00	79.00
	25-50% Percentile	10.00	0.00	64.50	3.24	61.00	64.50	69.00
	50-75% Percentile	12.00	0.00	69.58	8.02	58.00	69.00	86.00
	75-100% Percentile	13.00	0.00	67.92	4.73	61.00	68.00	78.00
Bone Scan Score	Overall	48.00	2.00	2.52	0.68	1.00	3.00	3.00
	0-25% Percentile	15.00	0.00	2.13	0.92	1.00	2.00	3.00
	25-50% Percentile	9.00	1.00	2.67	0.50	2.00	3.00	3.00
	50-75% Percentile	11.00	1.00	2.55	0.52	2.00	3.00	3.00
	75-100% Percentile	13.00	0.00	2.85	0.38	2.00	3.00	3.00
Grade	Overall	41.00	9.00	2.15	0.79	1.00	2.00	3.00
	0-25% Percentile	14.00	1.00	2.50	0.65	1.00	3.00	3.00
	25-50% Percentile	8.00	2.00	1.62	0.74	1.00	1.50	3.00
	50-75% Percentile	9.00	3.00	2.00	0.87	1.00	2.00	3.00
	75-100% Percentile	10.00	3.00	2.20	0.79	1.00	2.00	3.00
In Remission (Yes = 1)	Overall	50.00	0.00	0.28	0.45	0.00	0.00	1.00
	0-25% Percentile	15.00	0.00	0.67	0.49	0.00	1.00	1.00
	25-50% Percentile	10.00	0.00	0.20	0.42	0.00	0.00	1.00
	50-75% Percentile	12.00	0.00	0.17	0.39	0.00	0.00	1.00
	75-100% Percentile	13.00	0.00	0.00	0.00	0.00	0.00	0.00
Nadir PSA	Overall	50.00	0.00	16.36	39.25	0.10	0.95	183.00
	0-25% Percentile	15.00	0.00	0.19	0.03	0.10	0.20	0.20
	25-50% Percentile	10.00	0.00	0.65	0.20	0.30	0.70	0.90
	50-75% Percentile	12.00	0.00	3.30	2.53	1.00	1.95	8.00
	75-100% Percentile	13.00	0.00	59.15	59.97	10.00	38.00	183.00
Observation Time (Months)	Overall	50.00	0.00	28.46	18.39	1.00	28.00	75.00
	0-25% Percentile	15.00	0.00	40.00	12.57	24.00	40.00	75.00
	25-50% Percentile	10.00	0.00	32.60	19.40	9.00	31.00	60.00
	50-75% Percentile	12.00	0.00	29.00	19.40	6.00	27.50	60.00
	75-100% Percentile	13.00	0.00	11.46	8.68	1.00	8.00	26.00
Pre-Treatment PSA	Overall	43.00	7.00	670.75	1287.64	4.80	127.00	4797.00
	0-25% Percentile	11.00	4.00	783.68	1305.63	5.00	127.00	3405.00
	25-50% Percentile	9.00	1.00	148.31	157.72	4.80	90.00	524.00
	50-75% Percentile	11.00	1.00	486.55	1154.87	10.00	96.00	3946.00
	75-100% Percentile	12.00	1.00	1127.92	1747.09	25.00	258.50	4797.00
Performance Status	Overall	48.00	2.00	80.83	11.08	50.00	80.00	100.00
	0-25% Percentile	15.00	0.00	83.33	12.34	50.00	80.00	100.00
	25-50% Percentile	9.00	1.00	84.44	8.82	70.00	80.00	100.00
	50-75% Percentile	11.00	1.00	78.18	13.28	50.00	80.00	90.00
	75-100% Percentile	13.00	0.00	77.69	8.32	60.00	80.00	90.00

Table 1: Descriptive statistics for the overall sample and within groups based on sample quartiles of nadir PSA

0. In Model 3, the intercept corresponds to the log odds of relapse within 24 months among subjects with nadir PSA of 0 ng/ml, a bone scan score of 1, and a performance score of 0.

3. (a) To assess the association between mean nadir PSA level and relapse within 24 months after adjustment for bone scan score and performance status, we will fit a linear regression model to the data using nadir PSA as the outcome and the indicator of relapse within 24 months as the predictor of interest. We will also adjust for bone scan score as a categorical variable and performance status as a continuous variable. Standard errors for coefficient estimates will be calculated using robust sandwich estimation, and asymptotic 95% confidence intervals will be calculated for each coefficient estimate. We will perform a hypothesis test at the 0.05 significance level of the hypothesis that the coefficient associated with relapse status is equal to zero. Rejection of this hypothesis will be taken as evidence of an association between mean nadir PSA level and relapse within 24 months after adjustment for bone scan score and performance status.

   Based on this regression analysis, we estimate that the difference in mean nadir PSA between those who have relapsed within 24 months and those still in remission after 24 months is 23.36 ng/ml, with those who have relapsed estimated to have greater mean nadir PSA. A 95% confidence interval for this estimate is (1.89, 44.83). Based on a P-value of 0.033, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between nadir PSA and relapse within 24 months.

   (b) To assess the association between geometric mean nadir PSA level and relapse within 24 months after adjustment for bone scan score and performance status, we will fit a linear regression model to the data using log nadir PSA as the outcome and the indicator of relapse within 24 months as the predictor of interest. We will also adjust for bone scan score as a categorical variable and performance status as a continuous variable. Standard errors for coefficient estimates will be calculated using robust sandwich estimation, and asymptotic 95% confidence intervals will be calculated for each coefficient estimate. We will perform a hypothesis test at the 0.05 significance level of the hypothesis that the coefficient associated with relapse status is equal to zero. Rejection of this hypothesis will be taken as evidence of an association between geometric mean nadir PSA level and relapse within 24 months after adjustment for bone scan score and performance status.

   Based on this regression analysis, we estimate that the difference in geometric mean nadir PSA between those who have relapsed within 24 months and those still in remission after 24 months is 13.73 ng/ml, with subjects who have relapsed estimated to have greater geometric mean nadir PSA. A 95% confidence interval for this estimate is (4.55, 41.45). Based on a P-value$<$ 0.001, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between geometric mean nadir PSA and relapse within 24 months.

4. (a) One merit of Model 1 over Model 2 is that keeping nadir PSA on its original scale may improve the interpretability of the results. However, Model 2 has the advantage that nadir PSA may be more scientifically meaningful when considered on the log scale, as is often the case with concentrations. One merit of the spline model used in Model 3 is that it allows for

flexible modeling of the association between nadir PSA and the log odds of relapse within 24 months. However, fitting a spline does typically require more data compared to other approaches, and may not perform well over ranges of the predictor of interest where the data are sparse.

One advantage of the linear regression approach used in part (a.) of Question 3 compared to the approach used in part (b.) is that people are generally more familiar with the arithmetic mean than the geometric mean. One advantage of the linear regression approach used in part (b.) of Question 3 is that the geometric mean is less susceptible to outliers than the arithmetic mean.

(b) We are interested in relapse within 24 months. However, our measurement of nadir PSA corresponds to the smallest observed PSA level over the entire course of follow-up, which could occur either before or after 24 months. In the first three analyses we considered where relapse within 24 months is the outcome, it is therefore possible for some subjects' nadir PSA to have occurred after 24 months. A similar issue exists for the two linear regression models considered in Question 3.