Exam 1 Review

A medical center urology group is studying the association between prostate-specific antigen (PSA) levels and several clinical measurements on 97 men with cancer about to undergo prostatectomies as treatment. We focus here on the association of PSA levels (this is a noninvasive blood test result) and Gleason scores (cancer severity score ranging from 6 to 8 with higher scores indicating worse prognosis). Note that PSA levels and Gleason scores are higher for this group than they would be for a random sample of men because all of these men have advanced cancers. The data are available in some form as a file beginning “prostate.xxx” on the course website.

There are 9 variables for each man:

id = identification number

psa = prostate-specific antigen level (mg/ml)

cancvol = estimate of prostate cancer volume (cc)

weight = prostate weight (gm)

age = age of patient (yrs.)

bph = amount of benign prostatic hyperplasia (cm2 )

sem = presence/absence of seminal vesicle invasion

capspen = degree of capsular penetration (cm)

gleason = grade of disease (6, 7, or 8 for these men)

1. What type of study is this (observational or experimental)? Explain your reasoning.

This is an observational study. The variables of interest are clinical measurements and are therefore not being acted upon by an experimental factor. Therefore, the study is interested in changes in these measurements under an unaltered state.

(b) Identify the following as specifically as possible:

i. population – men with cancer about to undergo prostatectomies

ii. sample – the 97 men in the study

iii. explanatory variable – Gleason scores

iv. response variable - PSA Levels

v. confounding variables – age, weight, cancer volume, SEM

(c) Before examining the results of an ANOVA, we should check our model assumptions. This can be done in two ways: by examining the data in each group for normality and constant variance or by running an ANOVA and examining the residuals. Here, it is sufficient to take the first approach. Comment on the assumptions for ANOVA.

|  |  |
| --- | --- |
| Assumption | Code |
| Normal Distribution – overall PSA shows significant right skew in a visual test. Breaking down by Gleason ordinals shows the same pattern. |  |
| Equal variances – A visual test for equal variances highlights significant differences between Gleason score 8 and the rest of the data. |  |
| Independence – no mention is made of any sort of randomization or description of sampling. We will assume independence. |  |

(d) A logarithmic or square root transformation often helps with continuous positive measurements such as the PSA measurements. Reconsider the assumptions using log transformed data and square root transformed data. Create new variables by including something similar to LOGY=LOG(Y) and SQRTY=SQRT(Y) in the DATA step of your SAS program (or equivalent in R).

i. Do the transformed data satisfy the model assumptions?

Log transformed

|  |  |
| --- | --- |
| Assumption | Code |
| Normal Distribution – log transformed overall PSA shows a relatively normal distribution in a visual test. Breaking down by Gleason ordinals shows the same pattern. |  |
| Equal variances – A visual test for equal variances shows relatively similar variances. It is possible this assumption could still be violated. |  |
| Independence – no mention is made of any sort of randomization or description of sampling. We will assume independence. |  |

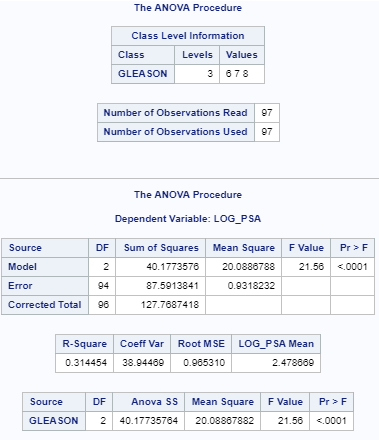
Square root transformation

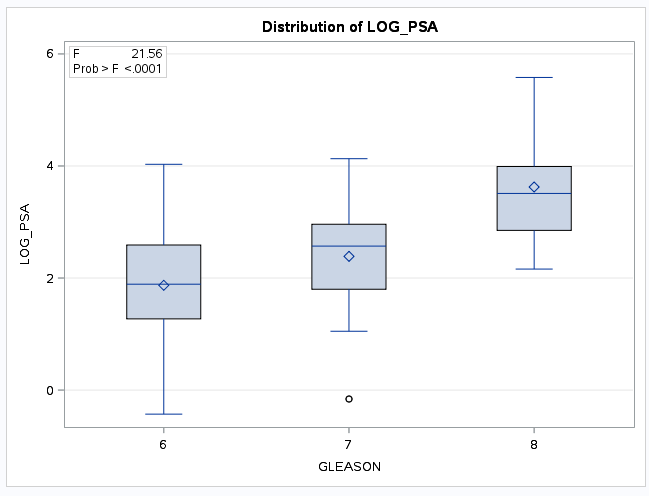
|  |  |
| --- | --- |
| Assumption | Code |
| Normal Distribution – square root overall PSA shows a right skewed dataset. Breaking down by Gleason ordinals shows the same pattern. |  |
| Equal variances – A visual test for equal variances shows that Gleason level 8 variances are still different from the other levels. |  |
| Independence – no mention is made of any sort of randomization or description of sampling. We will assume independence. |  |

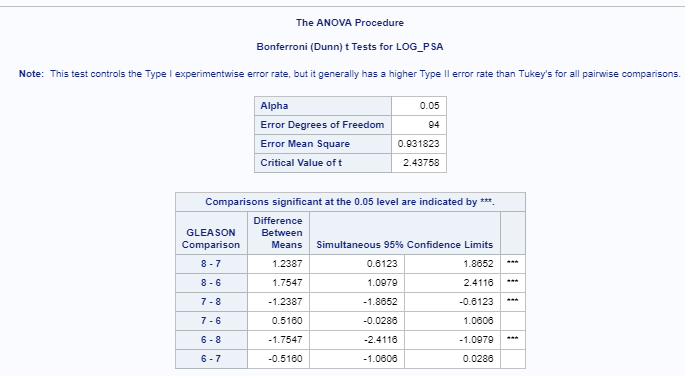
ii. Which transformation works best? Log transformation more closely aligns to ANOVA assumptions.

(e) For the remainder of the question, use the log transformed data. Is there evidence that median PSA levels vary across groups? You do not need to perform a complete analysis, but information you would put in the concluding paragraph of a complete analysis should suffice.

(f) Identify the pairs of groups which differ significantly using the Bonferroni approach. You do not need to perform a complete analysis, but you should provide evidence to support your conclusion.







There is significant evidence (p < .0001) that means between the following Gleason scores differ at the 95% confidence level:

G8-G7: [.6123, 1.8652]

G8-G6: [1.0979, 2.4116]

G7-G6: [-1.8652, -.6123]

G8-G6: [-2.4116, -1.0979]

g) Define and analyze contrasts to address the following:

i. linear trend of PSA vs. Gleason scores.

ii. quadratic trend of PSA levels vs. Gleason scores.

iii. What does it mean if both contrasts are significant?

(h) Describe your findings on these data in a paragraph.