AMS 315
Data Analysis
Chapter Eight Lecture Notes
Inferences about More than Two Population Central Values
Spring 2023

Context

The procedures in this chapter generalize the test of the equality of means of two independent populations. This generalization is often called the one-way layout. While this design has somewhat limited value in practice, the material in this chapter is fundamental for further generalizations. The key ideas that are first developed in the one-way analysis of variance are: the generalization of the t-test, the expected mean square calculation (which is described in Chapter 14 and is crucial for power calculations), and the introduction to multiple testing of hypotheses in Chapter 9.

The Model of Observations in a Completely Randomized Design

The usual "effects" model is $Y_{ij} = \mu + \alpha_i + \sigma_{1W} Z_{ij}$, for $i = 1, \ldots, I$ (where *I* is the number of treatment settings), $j = 1, \ldots, J_i$, and $\sum_{i=1}^{I} J_i \alpha_i = 0$. The use of Z_{ij} in this model is the assumption that the dependent variable data is normally distributed and independent. The use of the multiplier σ_{1W} is the assumption that the variances within groups are homogeneous. The important assumption is independence of the error terms. This is guaranteed when there is a random assignment of experimental unit to treatments. Sometimes researchers apply these techniques to data not generated by a randomized experiment. In that event, checking the assumption of independence is crucial. The $\{\alpha_i\}$ parameters are called the treatment effects. Under the effects model, $E(Y_{ij}) = \mu + \alpha_i$, and the distribution of Y_{ij} is $NID(\mu + \alpha_i, \sigma_{1W}^2)$.

OLS Estimates

A model that is equivalent to the effects model is called the means model and is $Y_{ij} = \mu_i + \sigma_{1W} Z_{ij}$, where $\mu_i = \mu + \alpha_i$. The sum of squares function is then $SS(m_1, \ldots, m_I) = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (y_{ij} - m_i)^2$. We seek values of the arguments that make the *SS* function as small as possible. As before, we take the partial derivatives and solve the normal equations.

Partial derivatives

One must calculate in turn $\frac{\partial}{\partial m_1} SS(m_1,\ldots,m_I),\ldots,\frac{\partial}{\partial m_I} SS(m_1,\ldots,m_I)$. First, focus on $\frac{\partial}{\partial m_1} SS(m_1,\ldots,m_I)$:

$$\frac{\partial}{\partial m_1} SS(m_1,\ldots,m_I) = \frac{\partial}{\partial m_1}\sum_{i=1}^{I}\sum_{j=1}^{J_i}(y_{ij}-m_i)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\frac{\partial}{\partial m_1}(y_{ij}-m_i)^2 \text{. Now,}$$

$$\frac{\partial}{\partial m_1} SS(m_1,\ldots,m_I) = \sum_{j=1}^{J_1}\frac{\partial}{\partial m_1}(y_{1j}-m_1)^2 + \sum_{i=2}^{I}\sum_{j=1}^{J_i}\frac{\partial}{\partial m_1}(y_{ij}-m_i)^2 .$$

One has to be careful with the partial derivative calculations. For observations from the first treatment,

$\frac{\partial}{\partial m_1}(y_{1j}-m_1)^2 = 2(y_{1j}-m_1)(\frac{\partial}{\partial m_1}(y_{1j}-m_1)) = 2(y_{1j}-m_1)(-1)$. For observations from the second and other treatments,

$\frac{\partial}{\partial m_1}(y_{2j}-m_2)^2 = 2(y_{2j}-m_2)(\frac{\partial}{\partial m_1}(y_{2j}-m_2)) = 2(y_{2j}-m_2)(0) = 0$. That is,

$$\frac{\partial}{\partial m_1} SS(m_1,\ldots,m_I) = \sum_{j=1}^{J_1}[-2(y_{1j}-m_1)] + \sum_{i=2}^{I}\sum_{j=1}^{J_i}0 = -2\sum_{j=1}^{J_1}(y_{1j}-m_1) = -2[\sum_{j=1}^{J_1}y_{1j} - J_1 m_1].$$

In general, $\frac{\partial}{\partial m_i} SS(m_1,\ldots,m_I) = -2[\sum_{j=1}^{J_i}y_{ij} - J_i m_i],\ i=1,\ldots,I.$

Normal Equations

Let $(\hat{\mu}_1,\ldots\hat{\mu}_I)$ be one of the solutions to the normal equations. Then, the first normal equation is

$\frac{\partial}{\partial m_1} SS(\hat{\mu}_1,\ldots,\hat{\mu}_I) = -2[\sum_{j=1}^{J_1}y_{1j} - J_1\hat{\mu}_1] = 0$. This can easily be solved to obtain

$\sum_{j=1}^{J_1}y_{1j} - J_1\hat{\mu}_1 = 0$ or $\hat{\mu}_1 = \frac{\sum_{j=1}^{J_1}y_{1j}}{J_1} = \bar{y}_1 = y_{1\bullet}$. The same analysis holds for the other treatment settings so that $\hat{\mu}_i = \frac{\sum_{j=1}^{J_i}y_{ij}}{J_i} = \bar{y}_i = y_{i\bullet},\ i=1,\ldots,I.$

The treatment model $Y_{ij} = \mu + \alpha_i + \sigma_{1W} Z_{ij}$, $j = 1, \ldots, J_i$, and $\sum_{i=1}^{I} J_i \alpha_i = 0$ has $I+1$ parameters (namely $\mu, \alpha_1, \ldots, \alpha_I$). The constraint on the treatment effects that $\sum_{i=1}^{I} J_i \alpha_i = 0$ is needed to make the parameters of the model and hence the OLS estimates unique. The OLS estimates are $\hat{\mu} = \frac{\sum_{i=1}^{I} J_i \hat{\mu}_i}{\sum_{i=1}^{I} J_i} = \frac{\sum_{i=1}^{I} J_i y_{i\bullet}}{\sum_{i=1}^{I} J_i} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i} y_{ij}}{\sum_{i=1}^{I} J_i} = y_{\bullet\bullet}$, where $\hat{\mu} = y_{\bullet\bullet}$ is called the grand mean (or overall mean) of the observations. Then,
$\hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} = y_{i\bullet} - y_{\bullet\bullet}$, $i = 1, \ldots, I$.

Sum of Squared Errors

As in Chapters 11 and 12, the minimized value of the *SS* function is the sum of squared error and is crucial for our analysis. Now,

$\min[\,SS(m_1, \ldots, m_I)] = SS(\hat{\mu}_1, \ldots, \hat{\mu}_I) = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (y_{ij} - y_{i\bullet})^2 = \sum_{i=1}^{I} (J_i - 1)s_i^2$, where

$s_i^2 = \frac{\sum_{j=1}^{J_i} (y_{ij} - y_{i\bullet})^2}{J_i - 1}$ is the usual sample variance estimator applied to the J_i observations from the *i*th setting of the treatment. Then the sum of squared error *SSE* is given by

$\min[\,SS(m_1, \ldots, m_I)] = SSE = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (y_{ij} - y_{i\bullet})^2 = \sum_{i=1}^{I} (J_i - 1)s_i^2$, with

$\sum_{i=1}^{I} (J_i - 1) = n - I$ degrees of freedom, where *n* is the total number of observations in the study.

Fisher's decomposition of the total sum of squares

I will now shift the discussion from a realized experiment to a planned experiment. That is, I will use the random variable notation. The total sum of squares is always

$SS_{Total} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - Y_{\bullet\bullet})^2$, with $n - 1 = \sum_{i=1}^{I} J_i - 1$ degrees of freedom. Then

$SS_{Total} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - Y_{\bullet\bullet})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - Y_{i\bullet} + Y_{i\bullet} - Y_{\bullet\bullet})^2$ and

$$SS_{Total} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - Y_{i\bullet})^2 + \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{i\bullet} - Y_{\bullet\bullet})^2 + 2\sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - Y_{i\bullet})(Y_{i\bullet} - Y_{\bullet\bullet}).$$

Recall that $SSE = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij} - Y_{i\bullet})^2$. Further,

$\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{i\bullet} - Y_{\bullet\bullet})^2 = \sum_{i=1}^{I}(Y_{i\bullet} - Y_{\bullet\bullet})^2\sum_{j=1}^{J_i}1 = \sum_{i=1}^{I}J_i(Y_{i\bullet} - Y_{\bullet\bullet})^2 = SS_{Treatment}$.The sum of squares due to treatment settings is defined to be

$\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{i\bullet} - Y_{\bullet\bullet})^2 = \sum_{i=1}^{I}J_i(Y_{i\bullet} - Y_{\bullet\bullet})^2 = SS_{Treatment}$ and has $I-1$ degrees of freedom.

Finally, $2\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij} - Y_{i\bullet})(Y_{i\bullet} - Y_{\bullet\bullet}) = 2\sum_{i=1}^{I}(Y_{i\bullet} - Y_{\bullet\bullet})\sum_{j=1}^{J_i}(Y_{ij} - Y_{i\bullet})$.

Since $\sum_{j=1}^{J_i}(Y_{ij} - Y_{i\bullet}) = 0,$ the cross-product term is 0. This proves that

$SS_{Total} = SSE + SS_{Treatment}$.

Analysis of Variance Table

These results are conventionally displayed in an analysis of variance table

Analysis of Variance Table
Complete Randomized Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatment	$I-1$	$\sum_{i=1}^{I}J_i(Y_{i\bullet} - Y_{\bullet\bullet})^2$	$SS_{Treatment}/(I-1)$	$\frac{MS_{Treatment}}{MSE}$
Error	$n-I$	$\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij} - Y_{i\bullet})^2 = \sum_{i=1}^{I}(J_i - 1)S_i^2$	$SSE/(n-I)$	
Total	$n-1$	$\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij} - Y_{\bullet\bullet})^2$		

As in Chapters 11 and 12, the statistical estimate of the variance parameter in the model is the mean squared error. The model is $Y_{ij} = \mu + \alpha_i + \sigma_{1W}Z_{ij}$, for $i = 1,\ldots,I$ (where *I* is the number of treatment settings), $j = 1,\ldots,J_i$, and $\sum_{i=1}^{I}J_i\alpha_i = 0$. Then $\hat{\sigma}_{1W}^2 = MSE$.

Tests of hypotheses

The most common issue is whether the expected value of the outcome variable is the same for each setting of the treatment. The usual null hypothesis is then

$H_0: \mu_1 = \mu_2 = \cdots = \mu_I$. The alternative hypothesis is $H_1: \mu_i \neq \mu_{i'}, i \neq i'$; that is, there is at least one pair of treatment settings with unequal means. An equivalent statement using the effects model is that $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$. The equivalent alternative hypothesis is $H_1: \alpha_i \neq \alpha_{i'}, i \neq i'$ for at least one pair of settings. The test statistic for this null hypothesis is $F = \dfrac{MS_{Treatment}}{MSE}$. The distribution of the test statistic under $H_0: \mu_1 = \mu_2 = \cdots = \mu_I$ is a central F distribution with $I-1$ numerator and $n-I$ denominator degrees of freedom. I call this test the "overall F-test" or "global F-test."

When the null hypothesis is true, the test statistic should be one, modulo statistical variability. When the alternative hypothesis is true, the test statistic should be greater than one modulo statistical variability. That is, the test of the null hypothesis is a right sided test.

When the global null hypothesis is rejected, researchers want to know which settings of the treatment variable are associated with larger expected values and which are associated with smaller expected values. Such questions lead to the issues of multiple comparisons, which is covered more deeply in Chapter Nine. A relatively simple approach is to use Fisher's protected t confidence intervals (which is also called Fisher's Least Significant Difference). Fisher's protected confidence intervals are calculated only when the global null hypothesis is rejected. Then one calculates a confidence interval for $E(Y_{ij}) - E(Y_{i'j'}) = \mu_i - \mu_{i'}$ for each pair of treatment settings (i, i') using a procedure analogous to the procedures in Chapter 6. These comparisons are called *post hoc* comparisons. For example, the 99% confidence interval for $\mu_1 - \mu_2$ would be $y_{1\bullet} - y_{2\bullet} \pm t_{2.576, n-I}\sqrt{MSE(\frac{1}{J_1} + \frac{1}{J_2})}$. When the experiment is underpowered, it might well happen that the global null hypothesis is rejected with no protected confidence interval excluding zero.

Balanced one-way layouts

Typically, the questions that I ask on examinations used a balanced one-way layout. That is, $J_1 = J_2 = \cdots = J_I = J$. This simplifies the calculations and permits natural and more complex issues discussed in Chapter Nine. The interpretation of the global F-test is more clear for balanced one-way layouts. There,

$$SS_{Treatments} = \sum_{i=1}^{I} J_i (Y_{i\bullet} - Y_{\bullet\bullet})^2 = \sum_{i=1}^{I} J (Y_{i\bullet} - Y_{\bullet\bullet})^2 = (I-1)J\left[\frac{\sum_{i=1}^{I}(Y_{i\bullet} - Y_{\bullet\bullet})^2}{I-1}\right].$$

The term in the square bracket is the usual unbiased estimator of the variance applied to the *I* values $\{Y_{1\bullet}, Y_{2\bullet}, \ldots, Y_{I\bullet}\}$. Then, under the global null hypothesis

$$E_0\left[\frac{\sum_{i=1}^{I}(Y_{i\bullet}-Y_{\bullet\bullet})^2}{I-1}\right]=\mathrm{var}(Y_{i\bullet})=\frac{\sigma_{1W}^2}{J} \text{ so that } E_0(SS_{Treatments})=(I-1)J\left[\frac{\sigma_{1W}^2}{J}\right]=(I-1)\sigma_{1W}^2 .$$

That is, under the global null hypothesis, $E_0(MS_{Treatments})=E(MSE)$. The global F-test is then the ratio of two quadratic forms with the same expectation under the global null hypothesis. When the global null hypothesis does not hold, $E(MS_{Treatments})>E(MSE)$. One should reject the global null hypothesis when the global F-test is larger than one.

Example Past Examination Questions

1. A research team wishes to specify a manufacturing process so that Y, the area in a product affected by surface flaws is as small as possible. They have four levels of concentration of a chemical used to wash the product before the final manufacturing step and want to determine whether the concentration level causes a change in $E(Y)$. They run a balanced one-way layout with 6 observations for each concentration with level 1 set at 10%, level 2 set at 15%, level 3 set at 20%, and level 4 set at 25%. They run a balanced one-way layout with 6 observations for each treatment. They observe that $y_{1.}=264.5,\ y_{2.}=255.9,\ y_{3.}=216.2, \text{ and } y_{4.}=263.8,$ where $y_{i.}$ is the average of the observations taken on the ith level. They also observe that $s_1^2=411.9,\ s_2^2=522.2,\ s_3^2=631.8, \text{ and } s_4^2=521.9,$ where s_i^2 is the unbiased estimate of the variance for the observations taken on the ith level.

a. Complete the analysis of variance table for these results; that is, be sure to specify the degrees of freedom, sum of squares, mean square, and F-test.
b. What is your conclusion? Use significance levels set to 0.10, 0.05, and 0.01. Make sure that you discuss the optimal setting of the concentration level and how you could document it.
c. In examinations, I add parts asking for the decomposition into linear, quadratic, and cubic components. See Chapter Nine Problems.

Answers:

Analysis of Variance Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	
Treatments	9467.4	3	3155.8	F=6.046
Error	10439.0	20	521.95	
Total	19906.4	23		

Answer: The critical values for the F-test are 2.38 for the 0.10 level, 3.10 for the 0.05 level, and 4.94 for the 0.01 level. Reject the null hypothesis that the mean responses of the four treatments are equal. The optimum setting is to use 20% as the concentration setting. This can be confirmed with Fisher's protected confidence intervals.

8.4 Checking on the AOV Conditions

The most important assumption is that of independence. This is guaranteed in a randomized experiment in which the experimental units are randomly assigned to treatment. When the data do not come from a randomize experiment, this assumption should be checked carefully. Common problems occur when time series data (for example, an exchange rate on successive days as the dependent variable) is used. Also data describing a geographical area such as a census tract have spatial autocorrelation. Data on students from the same class will be correlated because of the common instruction.

The analysis procedures for balanced analyses of variances are not sensitive to violations of the normality assumption and the homogeneity of variance assumption. The residuals in a one-way AOV are $r_{ij} = y_{ij} - y_{i\bullet}$. Residual analysis is simple for this model. One can and should generate a probability plot of the residuals. Closeness of the plot to a straight suggests that the assumption of normality appears to be true. Hartley's $F_{\max} = \dfrac{S^2_{\max}}{S^2_{\min}}$ is sensitive to normality. The Brown-Forsythe-Levene test for homogeneity of variance is more robust to violations of the assumption of normality. Many statistical packages will calculate this test, and you should use it routinely. There is another more robust test of this null hypothesis that corrects for the estimated kurtosis of the sampled random variables. Some statistical packages report this test as well or instead of Levene's test. If the hypothesis of constant variance is rejected, there are two common next steps. One is to use weighted least squares (with weights reflecting the difference in variance of observations), and the other approach is to transform the data to lessen the differences in variance. These transformations are called variance stabilizing transformations and are commonly used. They are helpful, especially when predictions of future values are to be made.

8.5 An Alternative Analysis: Transformations of the Data

Lecture Material on the "Delta Method":

The objective is to calculate the approximate mean and variance of a random variable $W = f(Y)$. The random variable Y has expected value μ_Y and variance σ_Y^2, and the function f has finite derivatives. The delta method approximates the value of W using the first term of the Taylor series: $W \cong f(\mu_Y) + f'(\mu_Y)(Y - \mu_Y)$. Then, $E(W) \cong E[f(\mu_Y) + f'(\mu_Y)(Y - \mu_Y)] = E[f(\mu_Y)] + E[f'(\mu_Y)(Y - \mu_Y)].$ Now $E[f(\mu_Y)] = f(\mu_Y),$ because $f(\mu_Y)$ is a constant. For the second term, $E[f'(\mu_Y)(Y - \mu_Y)] = f'(\mu_Y)E[(Y - \mu_Y)] = 0.$ The conclusion is that $E(W) \cong f(\mu_Y)$. The result that $E[F(r)] \cong F(\rho)$ is an application of this result.

The deviation $W - E(W) \cong f(\mu_Y) + f'(\mu_Y)(Y - \mu_Y) - f(\mu_Y) = f'(\mu_Y)(Y - \mu_Y).$ Then $E\{[W - E(W)]^2\} \cong E\{[f'(\mu_Y)(Y - \mu_Y)]^2\} = [f'(\mu_Y)]^2 E[(Y - \mu_Y)^2].$

That is, $\mathrm{var}(W) \cong [f'(\mu_Y)]^2 \,\mathrm{var}(Y).$

Example Problem

The random variable Y has the Poisson distribution with expected value μ. Find the approximate mean and variance of $\sqrt{Y}.$

Solution

The expectation calculation is easy: $E(\sqrt{Y}) \cong \sqrt{\mu}.$ Next, the first derivative calculation: $f(y) = y^{0.5}$ so that $f'(y) = 0.5y^{-0.5}.$ From Chapter 4, $\mathrm{var}(Y) = \mu.$ Finally, $\mathrm{var}(\sqrt{Y}) \cong [f'(\mu_Y)]^2 \,\mathrm{var}(Y) = [0.5\mu^{-0.5}]^2 \mu = 0.25.$

Comment

Since $\mathrm{var}(\sqrt{Y}) \cong 0.25$ independently of the expected value of the Poisson random variable, the square root transformation is said to be a variance stabilizing transformation for a Poisson random variable. Another transformation of a Poisson random variable is $W = \sqrt{Y} + \sqrt{Y+1}$ has approximate expected value $E(W) \cong \sqrt{\mu} + \sqrt{\mu+1}$ and approximate variance $\mathrm{var}(W) \cong 1$. This transformation is an example of a Freeman-Tukey deviate.

Example examination problem

The random variable Y, $Y > 0$, has $E(Y) = \theta$ and $\mathrm{var}(Y) = \theta^3, \theta > 0$. Find the approximate mean and variance of $W = \ln(Y)$.

Answer: $E(W) \cong \ln(\theta),$ and $\mathrm{var}(W) \cong [f'(\theta)]^2 \,\mathrm{var}(Y) = [1/\theta]^2 \theta^3 = \theta.$

Exploratory Data Analysis Tool to Identify Variance Stabilizing Transformation

When the dependent variable in an analysis of variance is always positive, calculate $y_{i\bullet}$ and $s_i^2 = \dfrac{\sum_{j=1}^{n_i} (y_{ij} - y_{i\bullet})^2}{n_i - 1}$, where i indexes the treatments in the analysis of variance. Plot $\log(s_i)$ against $\log(y_{i\bullet})$ and fit a straight line to the data. Call the slope *m*. When $m \cong 0$, no transformation is necessary. When $m \neq 0$, then analyze the transformed values $t_{ij} = y_{ij}^{1-m}$, with $m \neq 1$. When $m = 1$, use $t_{ij} = \log(y_{ij})$. There is a related set of techniques called the Box-Cox transformations that is also helpful.

Probability theory behind tool

The random variable Y has mean μ_Y and standard deviation σ_Y such that $\ln(\sigma_Y) = a + m\ln(\mu_Y)$. Then $\sigma_Y = \exp[\ln(\sigma_Y)] = \exp[a + m\ln(\mu_Y)] = \exp[a + \ln(\mu_Y^m)]$.
Further,
$\sigma_Y = \exp[\ln(\sigma_Y)] = \exp[a + m\ln(\mu_Y)] = \exp[a + \ln(\mu_Y^m)]$. This means that the standard deviation of Y is related to the expected value of Y in the equation $\sigma_Y = \exp[a + \ln(\mu_Y^m)] = \exp(a) \times \exp[\ln(\mu_Y^m)] = c\mu_Y^m$. For $W = f(Y)$, $\text{var}(W) \cong [f'(\mu_Y)]^2 \text{var}(Y)$, which is equivalent to $\sigma_W \cong | f'(\mu_Y) | \sigma_Y$. For functions f that are monotonically increasing, $\sigma_W \cong f'(\mu_Y)\sigma_Y$.

We seek to choose f so that $\sigma_W \cong f'(\mu_Y)\sigma_Y = k$, where k is a constant. That is, we seek f so that $\sigma_W \cong f'(\mu_Y)c\mu_Y^m = k$. That is, we seek f such that $f'(\mu_Y) = (\frac{k}{c})\mu_Y^{-m}$. For $m \neq 1$, anti-differentiation finds the function to be $f(\mu_Y) = c'\mu_Y^{1-m} + c''$. Typically, $c' = 1$ and $c'' = 0$. The required transformation of Y is $W = f(Y) = Y^{1-m}$, $m \neq 1$. For $m = 1$, $W = f(Y) = \ln(Y)$.

Example test problem

The random variable Y has $E(Y) = \theta$ and $\text{var}(Y) = \theta^{2.5}, \theta > 0$. Find the transformation W that makes the variance of W approximately constant. What are the approximate mean and variance of W?

Solution: Since $\sigma_Y = \theta^{1.25}$, $\ln(\sigma_Y) = 1.25\ln(\theta) = 1.25\ln(\mu_Y)$. The rule for finding a variance stabilizing transformation is to use the transformation $W = Y^{1-1.25} = \frac{1}{Y^{0.25}}$. From the delta method, $E(W) \cong \frac{1}{\theta^{0.25}}$. The derivative of the transformation function is $f'(y) = -0.25y^{-1.25}$. Then $\text{var}(W) \cong (-0.25 \times \theta^{-1.25})^2 \theta^{2.5} = (0.25)^2$.