(18) 1. Suppose that $W \sim \chi_3^2$ and that the distribution of X given $W = w$ is U[0, w] (that is uniform on 0 to w).

(5) (a) Find $E(X)$.

(8) (b) Find $Var(X)$.

(5) (c) Suppose $W = 4$. What is the value of the minimum mean square error forecast of X?

(10) 2. $X_i \sim U[0,1]$ for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^{n} X_i$. Suppose that $n = 100$. Use an appropriate approximation to estimate the value of $P(Y \leq 52)$.

(22) 3. $X_i \sim \text{iidN}(0, \sigma^2)$ for $i = 1, \ldots, 10$. A researcher is interested in the hypothesis H_o: $\sigma^2 = 1$ versus H_a: $\sigma^2 > 1$. He decides to reject the null if $\frac{1}{10}\sum_{i=1}^{10} X_i^2 > 1.83$.

(6) (a) What is the size of his test?

(6) (b) Suppose that alternative is true and that $\sigma^2 = 1.14$. What is the power of his test? .

(10) (c). Is the researcher using the best test? Explain.

(10) 4. Y_i is iid N(μ, σ^2) for i = 1, … , n. Let $\overline{Y}$ denote the sample mean, and consider $\hat{\mu} = \overline{Y} + 2$ as an estimator of μ. Suppose that the loss function is $L(\hat{\mu}, \mu) = (\hat{\mu} - \mu)^2$. Show that $\hat{\mu}$ is an inadmissible estimator of μ.

Problem 5 (17 points)

The table below shows the output from running a standard wage regression using a "generic" software package.

Dependent Variable: LNWAGE
Method: Least Squares
Date: 01/14/05 Time: 10:48
Sample: 1 550
Included observations: 550

Variable	Coefficient	Std. Error	t-Statistic
C	0.488168	0.098323	4.964941
FE	-0.306048	0.034442	-8.886016
UNION	0.207143	0.036850	5.621193
HISP	-0.027148	0.068828	
NONWH	-0.157265	0.054961	-2.861377
ED	0.074610		11.21589
EX	0.026191	0.004717	5.552082
EXSQ	-0.000308	0.000102	-3.034842

R-squared		Mean dependent var	1.681002
Adjusted R-squared		S.D. dependent var	0.490157
S.E. of regression	0.384520	Akaike info criterion	0.940795
Sum squared resid	80.13758	Schwarz criterion	1.003484
Log likelihood	-250.7185	F-statistic	
Durbin-Watson stat	1.995403	Prob(F-statistic)	

(a) *(4 points)* Find the t–Statistic for the coefficient on the explanatory variable HISP and the standard error for the coefficient on ED.

(b) *(3 points)* Perform a test of whether the coefficient of HISP equals 0 against the alternative that it is negative. Test at a 5% level of significance.

(c) *(6 points)* Find R^2.

(d) *(4 points)* Find the F–statistic for testing whether all the explanatory variables (excluding the constant) have coefficients equal to 0 (Hint: This is the statistic for the version of the F–test that assumes homeskedasticity).

Problem 6 (15 points)

Somebody regresses a variable, y, on two variable x_1 and x_2 (and a constant) and finds an R^2 of 0.07. If the sample size is 1200, do you conclude that y is unrelated to (x_1, x_2)? Discuss the assumptions underlying your answer.

Problem 7 (14 points)

Suppose that

$$y_i = x_i\beta + \varepsilon_i$$

where x_i is one–dimensional with $E[x_i] = E[\varepsilon_i] = 0$, $E[x_i^2] = E[\varepsilon_i^2] = 1$ and $E[x_i\varepsilon_i] = \rho$. Let $\widehat{\beta}_{ols}$ be the OLS estimator in a regression of y_i on x_i. Find the probability limit of $\widehat{\beta}_{ols}$. How does it depend on ρ?

Problem 8 (14 points)

Suppose that you estimate a linear regression model using families from a (large) sample of villages,

$$y_{ij} = x_{ij}'\beta + \varepsilon_{ij}$$

where $i = 1, \ldots, n$ denotes a village, and j denotes family. For simplicity assume that the sample consists of three families per village, so $j = 1, 2, 3$.. Assume that observations are independent and identically distributed across villages (so $\{(x_{ij}, \varepsilon_{ij})\}_{j=1}^{3}$ has the same distribution for all i, and $\{(x_{ij}, \varepsilon_{ij})\}_{j=1}^{3}$ is independent of $\{(x_{\ell j}, \varepsilon_{\ell j})\}_{j=1}^{3}$ provided that $i \neq \ell$). Also assume that the model is correctly specified in the sense that $E[\varepsilon_{ij} | x_{i1}, x_{i2}, x_{i3}] = 0$ for $j = 1, 2, 3$.

It is often unreasonable to assume that observations within a village are independent. This implies that the standard theory for OLS is invalid, and one cannot rely on the output from standard OLS packages for inference.

What is the asymptotic distribution (as n increases) of the OLS estimator? State any additional assumptions that you make.