STAT 371 Final Exam Summary

Statistics for Finance I

1 OLS and $R\beta$

- **Log-log** model: $\ln Y_t = \beta_1 + \beta_2 \ln X_t$, **semi-log** model: $\ln Y_t = \beta_1 + \beta_2 X_t$, **linear** model: $Y_t = \beta_1 + \beta_2 X_t$
- $\beta, \hat{\beta}$ is $k \times 1$, $Y, \hat{Y}$ is $n \times k$, X is $n \times k$, $\hat{U}, U$ is $n \times k$

Basic GLRM Framework:

- $\hat{\beta}_{OLS} = (X^t X)^{-1} X^t Y$
- $Var[U] = \sigma_U^2 I$
- $Var[\hat{\beta}] = \hat{\sigma}_u^2 (X^t X)^{-1} = \left[\frac{RSS}{n-k}\right] (X^t X)^{-1}$

$R\beta$ **Framework:**

- $H_0 : R\beta = r, H_1 : R\beta \neq r, q := \text{rank}(R)$
- $\hat{\beta}_R = \hat{\beta} + (X^t X)^{-1} R^t \left(R(X^t X)^{-1} R^t\right)^{-1} (r - R\hat{\beta})$
- $\hat{U}_R^t \hat{U}_R = y^t y^t - \hat{\beta}_R x^t y^t$
- $Var[\hat{\beta}_R] = \hat{\sigma}_u^2 (I - AR)(x^t x)^{-1} (I - AR)^t$ where $A = (x^t x)^{-1} R^t \left[R(x^t x)^{-1} R^t\right]^{-1}$
- We have the following equivalent statements

$$\begin{aligned} TSS &= RSS + ESS \\ Y^t Y - n\bar{Y}^2 &= \hat{U}^t \hat{U} + \hat{\beta}^t X^t Y - n\bar{Y}^2 \\ y^t y &= \hat{U}^t \hat{U} + \hat{\beta}^t x^t y \end{aligned}$$

Key Statistics:

- $R^2 = \frac{ESS}{TSS}$, $\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k}$
- $t = \frac{\hat{\beta}_1 - \beta_1}{sd(\hat{\beta}_1)} \sim t(n-k)$, $\frac{r - R\beta}{\sqrt{\hat{\sigma}^2 R (X^t X)^{-1} R^t}} \sim t(n-k)$ for $q = 1$
- $F_{Statistic} = \frac{ESS/(k-1)}{RSS/(n-k)} =\sim F(k-1, n-k)$ (ANOVA)
- We have for $q \geq 1$,

$$\begin{aligned} t^2 = F &= \frac{(R\hat{\beta} - r)^t [R(X^t X)^{-1} R^t](R\hat{\beta} - r)/q}{(\hat{U}^t \hat{U})/(n-k)} \\ &= \frac{(RSS_R - RSS_{UN})/q}{RSS_{UN}/(n-k)} \sim F(q, n-k) \end{aligned}$$

Special Matrices:

$X(X^t X)^{-1} X^t = \text{Proj}_X(\cdot)$

$M = (I - \text{Proj}_X) = (I - X(X^t X)^{-1} X^t) = \text{Proj}_{\hat{U}}(\cdot)$ where M is idempotent and of rank $n - k$

2 Model Selection and Specification

Problems with X:

1. Suppose that we have an **incorrect functional form** (p. 112).
   (a) Consequences?
      i. It *could* be *unbiased* and *inefficient*
      ii. The t and F tests are invalid
   (b) Detection?
      i. The *informal test* would be to just plot the data.
      ii. The *formal test* is the *Ramsey Reset test.*
2. Suppose that we are **underfitting.**
   (a) Let the true model be $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \mu_t$ but you omitted X_{3t} in the specification of your model. So you mistakenly specified $Y_t = \phi_1 + \phi_2 X_{2t} + v_t, v_t = \beta_3 X_{3t} + \mu_t$ and you get $E[v_t] = \beta_3 X_{3t} \neq 0$ and $Var[v_t] = \beta_3^2 Var(X_t) + \sigma_u^2 \neq c$ for a constant c.
   (b) Consequences?
      i. On the least square estimators, the OLS estimators are *biased* iff the excluded variable X_{3t} is correlated with the included variable X_{2t} ($r_{23} \neq 0$)
      ii. The t and F ratios are no longer valid.
   (c) Detection?
      i. An informal test is to add X_{3t} to your model and check if there is a change in R^2. If it goes up it is relevant.
      ii. Another informal test is to add X_{3t} to the model and check the changes in the new estimated coefficients. If there is a significant change, then we have a relevant variable.
      iii. The formal test is the Ramsey Reset test.
3. Suppose that we are **overfitting.**
   (a) Let the true model be $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$ but the mis-specified model be $Y_t = \theta_1 + \theta_2 X_{2t} + \theta_3 X_{3t} + v_t$ where X_{3t} is an irrelevant variable.
   (b) Consequences?
      i. The least squares estimator of the mis-specified model are *unbiased* and *consistent* but no longer *efficient*.

ii. The t and F ratios are no longer valid.

(c) Detection?

i. The informal tests are the same as above in the case of undefitting. However, $\bar{R}^2$ and the estimated coefficients are note expected to change very much.

ii. The more formal test is to test the restriction that $\theta_3 = 0$ using either the t test, the F test or the $t^2 = F$ statistic.

Ramsey Reset Test:

This is used to test for an incorrect functional form or for underfitting.

1. Run OLS and obtain $\hat{Y}_t$ and $\hat{Y}_t$ will incorporate the true functional form or the underfitting (if any exists)

2. Take the unrestricted model

$$Y_t = \phi_0 + \phi_1 X_t + \phi_2 \hat{Y}_t^2 + \phi_3 \hat{Y}_t^3 + ... + \phi_k \hat{Y}_t^k$$

and use the hypotheses $H_0 : \forall k, \phi_k = 0, H_1 : \exists k, \phi_k \neq 0$. **Usually** $k = 3$.

3. Compute

$$F = \frac{(RSS_R - RSS_{UN})/q}{RSS_{UN}/(n-k)} \sim F_{q,n-k}$$

and reject or don't reject H_0. If we don't reject then we have an incorrect functional form.

Errors in Y:

- We then have the equation $Y_t = \beta_1 + \beta_2 X_{2t} + \left(\underbrace{u_t + \xi_t}_{\epsilon_t} \right)$ where we call ϵ_t **composite error**.
- The least squares estimators in Y_t from above will remain unbiased but no longer efficient (see **proof** in notes; may be on the final exam)

Errors in X:

- We have the equation

$$Y = (X - V)\beta + U = X\beta + (\underbrace{U - V\beta}_{=\epsilon}) = X\beta + \epsilon$$

- The $\hat{\beta}_{OLS}$ from above is going to be biased in small samples and inconsistent in large samples (see **proof** in notes; may be on the final exam)

(**Central Limit Theorem**) Suppose that we have $X_1, ..., X_n$ i.i.d. r.v.s with mean μ and variance σ^2. Then,

$$\lim_{n\to\infty} \bar{X} \sim N(\mu, \sigma^2/n) \implies \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, V)$$

Instrumental Variables:

- We need to find a matrix $Z_{n\times l}$, $l \geq k$ such that it satisfies certain properties. These are
- $E[Z^t U] = 0$
- $E[Z^t X] = \Sigma_{ZX}$
- We premultiply the observed model by Z^t to get:
  - $Z^t Y = Z^t X\beta + Z^t U$ and so $\hat{\beta}_{IV} = (X^t Z Z^t X)^{-1} X^t Z Z^t Y = (Z^t X)^{-1} Z^t Y$
  - If $l = k$, $p \lim_{n\to\infty} \hat{\beta} = \beta + \Sigma_{ZX} \cdot 0 = \beta$ (we need invertibility of Σ_{ZX})
- Read the notes to understand the various properties and **proofs**.
- The problems here are:
  1. The $X's$ are stochastic
  2. $E[X^t \epsilon] \neq 0$
  3. The errors $\epsilon's$ are no longer white noise? (They are. See **proof** in notes)

Two-stages Least Squares:

- If $l > k$, we do a procedure called the two-stages least squares (2SLS):
  1. Regress X on Z and obtain a matrix of fitted values $\hat{X}$ (Project X onto Z). That is

$$\hat{X} = Z(Z^t Z)^{-1} Z^t X$$

  2. Regress Y on $\hat{X}$ and obtain $\hat{\beta}_{2SLS}$ where $\hat{\beta}_{2SLS} = (\hat{X}^t \hat{X})^{-1} \hat{X}^t Y = (X^t \text{Proj}_Z X)^{-1} X^t \text{Proj}_Z Y$
  3. We can show that $\hat{\beta}_{2SLS} = \hat{\beta}_{IV}$. To do this, multiply by $(Z^t Z)(Z^t Z)^{-1}$ in the equation for $\hat{\beta}_{IV}$ to get

$$\begin{aligned} \hat{\beta}_{IV} &= (X^t Z(Z^t Z)^{-1} Z^t X)^{-1} X^t Z(Z^t Z)^{-1} Z^t Y \\ &= (X^t \text{Proj}_Z X)^{-1} X^t \text{Proj}_Z Y = \hat{\beta}_{2SLS} \end{aligned}$$

3 Non-Spherical Disturbances

When we have serial correlation and heteroskedasticity on the error terms, we call these error terms *non-spherical disturbances*. This is when we have a covariance matrix that is not diagonalized and and has non-zero entries on the off-diagonal elements.

Sources of Heteroskedasticity:

(1) Nature of Y_t **(2)** Mis-specification **(3)** Transformations **(4)** Varying coefficients

Mathematical Representation of σ_t^2:

(1) $\sigma_t^2 = \sigma^2 Z_t^h$ for some $h \neq 0$ **(2)** $\sigma_t^2 = \alpha_0 + \alpha_1 Z_t$ **(3)** $\sigma_t = \alpha_0 + \alpha_1 Z_t$ **(4)** $\sigma_t^2 = f(Z_1, Z_2, ..., Z_n)$

Testing for Heteroskedasticity

1. Park Test

   (a) Park specified $\sigma_t^2 = \sigma^2 X_t^\beta e^{v_t}$ for the model $Y_t = \beta_1 + \beta_2 X_t + u_t$.

   (b) From here, we linearize the above equation to get $\ln \sigma_t^2 = \ln \sigma^2 + \beta \ln X_t + v_t$. Since $\hat{u}_t$ is observed, it is a proxy for u_t and

   $$Var(\hat{u}_t) = E[(\hat{u}_t - 0)^2] = E[\hat{u}_t^2]$$

   we use $\ln \hat{u}_t$ as a proxy for $\ln u_t$. Our new equation is then

   $$\ln \hat{u}_t^2 = \ln \sigma^2 + \beta \ln X_t + v_t$$

   where we hope that v_t is white noise.

   (c) Test the hypothesis that $H_0 : \beta = 0$ using a t test and reject or not reject the null hypothesis. If we reject, then we have heteroskedasticity.

2. White Test

   (a) Let $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ and regress Y on the X's to get a series of $\hat{u}_t$

   (b) Run the auxiliary regression (stated in R formula notation) $\hat{u}_t^2 \sim (X_{2t} + X_{3t})^2 + X_{2t}^2 + X_{3t}^2$

   (c) Compute R^2 from the previous regression

   (d) White showed that asymptotically, the quantity $W = nR^2 \sim \chi^2(k-1)$ where k is the number of all the parameters in the auxiliary regression (here $k = 6$) If the test statistic is larger than the critical at $\alpha = 5\%, k-1$ then we have heteroskedasticity.

3. Of course we don't know which of the explanatory variables is causing this, but we have some remedies:

   (a) Test using the White procedure

   (b) Narrow it down to a specific variable (could be in the model) or outside the model (one unknown variable)

      i. If it is coming from one of the X's, we can: try to replace it with a proxy, try to replace it with a combination of variables, drop it, do some transformations

      ii. It is due to Z (outside of the model), then: you could have underfitting; raise your specification and try to include that missing relevant variable

4. What if you know the exact form of heteroskedasticity?

   (a) Use General Least Squares

      i. *Example.* Suppose that heteroskedasticity is due to X_{2t} and it is taking the following form:

      $$\sigma_t^2 = \sigma^2 X_{2t}^h, h = 2$$

      How can we correct for this problem? We use the method of Weighted Least Squares, also known as Generalized Least Squares (GLS)

         A. To do this, we want to "divide by the $\sqrt{}$ of whatever is causing the heteroskedasticity

         B. So let's transform our model as follows

         $$\frac{Y_t}{\sqrt{X_{2t}^2}} = \frac{\beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t}{\sqrt{X_{2t}^2}}$$

         We then get

         $$Var\left[\frac{u_t}{\sqrt{X_{2t}^2}}\right] = \frac{1}{X_{2t}^2} Var[u_t] = \sigma^2$$

         and this new model is homoskedastic.

Serial Correlation:

1. Problem: $Cov(u_t, u_s) \neq 0$ for $t \neq s$

2. Sources: P. 162-164 (will be on the final exam)

3. Mathematical Representation:

   (a) Let the true model be $Y_t = \beta_1 + \sum_{i=2}^n \beta_i X_{it} + u_t$ such that $E[u_t] = 0, Var(u_t) = \sigma^2$ and $Cov(u_s, u_t) \neq 0$

(b) We will only consider the AR(1) (autoregressive 1) process given by

$$u_t = \rho u_{t-1} + \xi_t$$

with $E[\xi_t] = 0$, $Var[\xi_t] = \sigma_\xi^2$, $Cov(\xi_t, \xi_s) = 0$ for $t \neq s$, and $|p| < 1$

(c) Remark that the conversion of this form into a general linear process through the use of forward recursion gives$u_t = \xi_t + \sum_{k=1}^{\infty} \xi_{t-k}\rho^k$. This implies that $E[u_t] = 0$, $Var[u_t] = \frac{\sigma_\xi^2}{1-\rho^2}$. We also get that $Cov(u_t, u_{t-s}) = \frac{\rho^s \sigma_\xi^2}{1-\rho^2}$

4. Test: Durbin-Watson (D-W) [applies only to AR(1)]:

(a) The $d-$statistic is $d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2} \approx 2(1-\hat{\rho})$ with $\hat{\rho} = \frac{\sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{n}\hat{u}_{t-1}^2}$ due to the fact that $\sum \hat{u}_{t-1}^2 \approx \sum \hat{u}_t^2$.

(b) Remark that if: $\rho = -1 \implies d = 4$, $\rho = 1 \implies d = 0$, $\rho = 0 \implies d = 2$

(c) According to Durbin and Watson, if $d \in (d_L, d_U)$ the test is inconclusive for $d_L, d_U \in (0, 2)$ and similarly for a symmetric reflection across $\rho = 2$ (this other interval is $(4 - d_U, 4 - d_L)$). Otherwise we make conclusions based on the proximity of d. Using this, we have several tests related to this.

i. Test for autocorrelation (p. 169):

A. $H_0 : \rho = 0$; no autocorrelation, $H_1 : \rho \neq 0$; there exists autocorrelation

B. Calculate $d \approx 2 - 2\hat{\rho}$ and use the d table to get d_L and d_U; use α and $df_1 = n$, $df_2 = k - 1$

C. Reject, not reject, or say the test is inconclusive

5. Remedies: GLS (Aitken 1936)

(a) Set up: $Y_t = \beta_1 + \sum \beta_k X_{kt} + u_t$, $u_t = \rho u_{t-1} + \xi_t$

(b) Apply D-W and if autocorrelation exists, correct using:

i. Use **GLS** if ρ is known:

A. Set up the equation (1) $Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2,t-1}) + ... + \xi_t$ since (2) $u_t = \rho u_{t-1} + \xi_t$ where ξ_t is white noise.

ii. **Cochrane-Orcutt Iterative Procedure** if ρ is not known:

A. Run OLS on (2) $Y_t = \beta_1 + ... + \beta_k X_{kt} + u_t$ and obtain a series of residuals $\hat{u}_t$

B. Compute $\hat{\rho}_1 = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_{t-1}^2}$

C. Use $\hat{\rho}_1$ for autocorrelation by applying GLS to get the estimated version of (1)

D. Apply D-W to (1)

E. If H_0 is accepted, then stop; if H_0 is rejected, go back to (2) using $Y_t - \rho Y_{t-1}$ as the new proxy for Y_t

F. Keep iterating until $\hat{\rho}_s \approx \hat{\rho}_{s-1}$ and H_0 is accepted

iii. Remark that the above Iterative Procedure doesn't also converge very well (it converges to a random walk) if $\rho \approx 1$

4 Maximum Likelihood Estimation

In MLE, we do the following:

1. Assume a distribution for Y
2. Define the pdf of y_i as $f_i(y_i|\theta)$ for each i
3. Find the joint pdf of the n realizations, assuming independence, with $f(Y|\theta) = \prod_{i=1}^{n} f_i(y_i|\theta)$
4. Define the likelihood function $L(\theta|Y) = f(Y|\theta) = \prod_{i=1}^{n} f_i(y_i|\theta)$
5. Take the log of L as $l(\theta|Y) = \log L(\theta|Y)$
6. Find θ through $\hat{\theta} = \text{argmax}_{\{\theta \in \Theta\}} l(\theta|Y)$

MLE and the GLRM:

- We define a few matrices:

  1. Score Matrix: $S(\theta) = \frac{\partial l}{\partial \theta}_{(k+1)\times 1} = 0_{(k+1)\times 1}$
  2. Hessian Matrix:

$$H(\theta) = \frac{\partial^2 l}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta \partial \beta'} & \frac{\partial^2 l}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 l}{\partial (\sigma^2)^2} \end{bmatrix}_{(k+1)\times(k+1)}$$

  3. Fisher Information Matrix: $I(\theta) = -E[H(\theta)]$

- Working in the GLRM framework (that is $Y = X\beta + U$), we will assume that $u_t \sim N(0, \sigma^2)$ for all t. The first order conditions give us

  1. $\hat{\beta}_{ML} = (X^t X)^{-1} X^t Y = \hat{\beta}_{OLS}$
  2. $\hat{\sigma}_{ML}^2 = \frac{\hat{U}^t \hat{U}}{n}$

- In terms of unbiased-ness:

  1. $\hat{\beta}_{ML} = \hat{\beta}_{OLS} \implies$ the estimate is unbiased for β
  2. $\hat{\sigma}_{ML} \neq \hat{\sigma}_{OLS} \implies \hat{\sigma}_{ML}$ is biased and $E[\hat{\sigma}_{ML}] = \left(\frac{n-k}{n}\right)\sigma^2$

- In terms of efficiency,

  1. $\hat{\beta}_{ML} = \hat{\beta}_{OLS} \implies Var[\hat{\beta}_{ML}] = Var[\hat{\beta}_{OLS}] = \sigma^2(X^tX)^{-1}$ and so our estimate is efficient
  2. $Var(\hat{\sigma}^2_{ML}) = \frac{n-k}{n}\left(\frac{2\sigma^4}{n}\right) \neq \sigma^2$ which means that it is inefficient and biased.

- In conclusion,

  1. In small samples, $\hat{\beta}_{ML}$ is unbiased and efficient. $\hat{\sigma}_{ML}$ is biased and inefficient.
  2. In large samples, it can be shown that both estimators are consistent and asymptotically normal (not shown in this course); that is, $\hat{\theta}_{ML}$ is a CAN (consistent and asymptotically normal) estimator.
  3. We can also show that they achieve the *Cramer-Rao lower bound* (**proof** will be on the final)

Asymptotic Test using ML (LR test):

Here LR test refers to the likelihood ratio test. The procedure is as follows:

1. Start with the unrestricted model:

   (a) $\hat{\theta}_{ML} = \begin{bmatrix} \hat{\beta}_{ML} = (X^tX)^{-1}X^tY \\ \hat{\sigma}^2_{ML} = \frac{\hat{U}^t\hat{U}}{n} \end{bmatrix}$ where $\hat{U}^t\hat{U} = y^ty - \hat{\beta}^tx^ty$

   (b) $L(\hat{\theta}_{ML}|Y) = (2\pi\hat{\sigma}^2_{ML})^{-\frac{n}{2}}e^{-\frac{n}{2}}$

2. Then do the same thing with the restricted model:

   (a) $\hat{\theta}_R = \begin{bmatrix} \hat{\beta}_R = \hat{\beta}_{ML} + (...) \\ \hat{\sigma}^2_R = \frac{\hat{U}^t_R\hat{U}_R}{n} \end{bmatrix}$ where $H_0 : r = R\beta$

   (b) $L(\hat{\theta}_R|Y) = (2\pi\hat{\sigma}^2_R)^{-\frac{n}{2}}e^{-\frac{n}{2}}$

3. The Likelihood ratio test uses the fact that

$$\begin{aligned} LRT_{Statistic} &= -2\left[\ln L(\hat{\theta}_R) - \ln(\hat{\theta}_{ML})\right] \\ &= -2\ln\left(\frac{L(\hat{\theta}_R)}{L(\hat{\theta}_{ML})}\right) \sim \chi^2(q) \end{aligned}$$

   where $H_0 : r = R\beta$, $H_1 : r \neq R\beta$, $LRT_{Critical} = (\alpha = 5\%, q)$. If $LRT_{Stat} > LRT_{Crit}$ then reject H_0.

   (a) Remark that $LRT_{Statistic}$ can also be re-written as

$$\begin{aligned} LRT_{Statistic} &= -2\ln\left(\frac{\hat{\sigma}^2_R}{\hat{\sigma}_{ML}}\right)^{-n/2} \\ &= \ln\left(\frac{\hat{\sigma}^2_R}{\hat{\sigma}_{ML}}\right) = -2\ln(\Lambda) \end{aligned}$$

(One computation related to the likelihood ratio will be on the final)

5 Basic Sampling Concepts

In sampling, we care about 3 characteristics of the population:

1. Population Total $t = \sum_{i=1}^N Y_i$
2. Population Mean: $\bar{Y} = \frac{1}{N}\sum_{i=1}^N Y_i = \frac{t}{N}$
3. Population Proportion: p

5.1 Simple Random Sampling (SRS)

In SRS,

1. We use $\bar{y}$ (the sample mean) to estimate $\bar{Y}$. That is, $\bar{y}$ is an estimator for $\bar{Y}$. Here, $\bar{y} = \frac{1}{n}\sum y_i$ and has the properties:

   (a) $E[\bar{y}] = \bar{Y}$

   (b) $Var[\bar{y}] = (1-f)\frac{S^2}{n}$ where S^2 is the true population variance. But S^2 is not known so we use the sample variance $s^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2$. Therefore, $\widehat{Var[\bar{y}]} = (1-f)\frac{s^2}{n}$.

2. Let's examine how we use the sample to estimate the population total. We know that $t = N\bar{Y}$ and since $\bar{y}$ is an estimator for $\bar{Y}$, we can use $\widehat{t} = N\bar{Y}$ which will be our estimator for t. It has the following properties:

   (a) $E[\hat{t}] = t$

   (b) $Var(\hat{t}) = N^2Var(\bar{y}) = N^2(1-f)\frac{s^2}{2}$

3. We skip the estimator, $\hat{p}$, for p.

(Assignment 4, Question 7) We are given $N = 6$, a population set $U_{\text{Index}} = \{1, 2, 3, 4, 5, 6\}$ with $Y_i = \{3, 4, 3, 4, 2, 2\}$.

a) We get that the population mean is $\bar{Y}_i = 3$ and the population variance is $s^2 = 0.8$.

b) The possible number of SRS's is $\binom{6}{3} = 20$

c) The probability of 1 SRS drawn is 1 over the number of possible SRS's. That is $\frac{1}{20}$.

d) The probability distribution of the sample mean is found as follows. We generate a list of all possible 3 element combinations from Y_i and the corresponding estimator values. Use this information to create the frequency distribution for the estimator. In this case, the mean has the following distribution:

$$P\left(\bar{y} = \frac{7}{3}\right) = \frac{2}{20}, P\left(\bar{y} = \frac{8}{3}\right) = \frac{4}{20}, P\left(\bar{y} = \frac{9}{3}\right) = \frac{8}{20},$$

$$P\left(\bar{y} = \frac{10}{3}\right) = \frac{4}{20}, P\left(\bar{y} = \frac{11}{3}\right) = \frac{2}{20}$$

and so $E[\bar{y}] = 3 = E[\bar{Y}]$ with $Var(\bar{y}) = \sum(y_i - \bar{y})^2 Pr_i = 0.133$.

5.2 Stratified Sampling

(Assignment 4 Question 8) We are given that

$$U_{\text{index}} = \{1, 2, 3, 4, 5, 6, 7, 8\}, Y_i = \{\underbrace{1, 2, 4, 8}_{N_1}, \underbrace{4, 7, 7, 7}_{N_2}\}$$

where N_1 and N_2 are the first and second stratums respectively. We want to take SRS's from from stratums:

a) SRS_1 of size $n_1 = 2$:

The number of possible SRS_1 is $\binom{4}{2} = 6$. We then have:

Sample No.	y_i	$P(s_i)$	$\bar{y}$	$\hat{t} = N_1\bar{y}$
1	$\{1, 2\}$	$1/6$	1.5	$4 \times 1.5 = 6$
2	$\{1, 4\}$	$1/6$	2.5	$4 \times 2.5 = 10$
3	$\{1, 8\}$	$1/6$	4.5	$4 \times 4.5 = 18$
4	$\{2, 4\}$	$1/6$	3	$4 \times 3 = 12$
5	$\{2, 8\}$	$1/6$	5	$4 \times 5 = 20$
6	$\{4, 8\}$	$1/6$	6	$4 \times 6 = 24$