Practice Exam for Midterm 1

1. Consider the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2), i = 1, 2, \ldots, n$$

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the least squares estimates of β_0 and β_1 respectively. Let $r_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$.

(a) Given that $s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, prove that $E(s_{xy}^2) = \beta_1^2 s_{xx}^2 + \sigma^2 s_{xx}$.

(b) Given that $s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$, prove that $E(s_{yy}) = (n-1)\sigma^2 + \beta_1^2 s_{xx}$.

(c) Given that $\sum_{i=1}^n r_i^2 = s_{yy} - \frac{s_{xy}^2}{s_{xx}}$, show that $E(\sum r_i^2) = (n-2)\sigma^2$. (This provides an alternative way to prove that $\frac{\sum r_i^2}{n-2}$ is an unbiased estimator for σ^2.)

2. A company builds mobile phones and wants to be able to estimate its overhead costs. The company took a sample of its departments and looked at each of their total overhead costs, **overhead**, and their direct labour costs, **labour**. The figure shows the data. Consider the model $overhead = \beta_0 + \beta_1 labour + \epsilon$ and the R output.

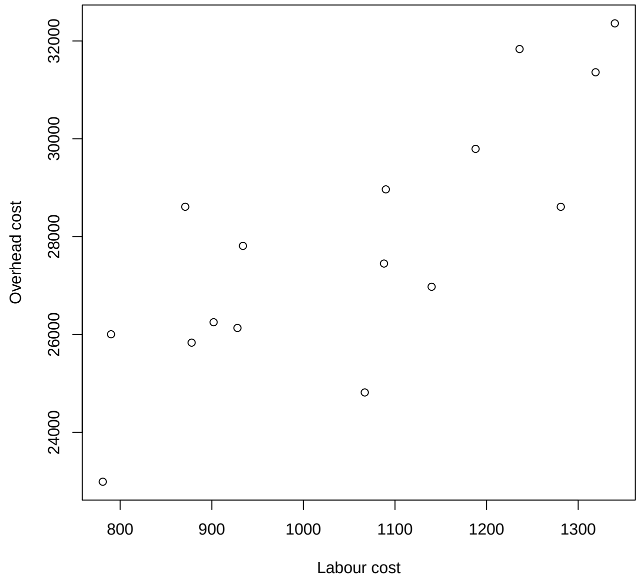

Figure 1: Data from cost study

```
> overhead.fit <- lm(overhead~labour)
> summary(overhead.fit)
        ...
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   16.310      2.421   6.737  NA
labour        10.982      2.268   4.843  NA
---


Residual standard error: 1.646 on 14 degrees of freedom
Multiple R-squared: 0.6262
```

(a) What are the numerical values of $\hat{\beta}_0$ and $\hat{\beta}_1$? Interpret $\hat{\beta}_1$.

(b) What is the numerical value of the standard error of the estimate of β_1?

(c) Compute a 95% confidence interval for β_1. t^* can be found in the attached t table.

(d) Test the hypothesis that $\beta_1 = 0$ at the significance level of 0.05.

(e) Compute numerically SSE and SSR for this data.

(f) Construct the ANOVA table for this model.

(g) Conduct an F test at the significance level of 0.05. State the hypotheses, test statistic and make your conclusion based on the critical value approach. The critical value can be found in the attached F table.

3. Suppose $Y = (y_1, y_2, y_3)'$ and $Y \sim MVN(\mu, \Sigma)$ with $\mu' = (1, 3, 0)$ and Σ is

$$\begin{pmatrix} 4 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 4 \end{pmatrix}$$

What is the distribution of $z = y_1 - 2y_2 - y_3$?

Upper percentile for a t_n distribution with n degrees of freedom

df/p	0.4	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.0005
1	0.325	0.727	1.38	3.08	6.31	12.7	31.8	63.7	637.0
2	0.289	0.617	1.06	1.89	2.92	4.30	6.96	9.92	31.6
3	0.277	0.584	0.978	1.64	2.35	3.18	4.54	5.84	12.9
4	0.271	0.569	0.941	1.53	2.13	2.78	3.75	4.60	8.61
5	0.267	0.559	0.920	1.48	2.02	2.57	3.36	4.03	6.87
6	0.265	0.553	0.906	1.44	1.94	2.45	3.14	3.71	5.96
7	0.263	0.549	0.896	1.41	1.89	2.36	3.00	3.50	5.41
8	0.262	0.546	0.889	1.40	1.86	2.31	2.90	3.36	5.04
9	0.261	0.543	0.883	1.38	1.83	2.26	2.82	3.25	4.78
10	0.260	0.542	0.879	1.37	1.81	2.23	2.76	3.17	4.59
11	0.260	0.540	0.876	1.36	1.80	2.20	2.72	3.11	4.44
12	0.259	0.539	0.873	1.36	1.78	2.18	2.68	3.05	4.32
13	0.259	0.538	0.870	1.35	1.77	2.16	2.65	3.01	4.22
14	0.258	0.537	0.868	1.35	1.76	2.14	2.62	2.98	4.14
15	0.258	0.536	0.866	1.34	1.75	2.13	2.60	2.95	4.07
16	0.258	0.535	0.865	1.34	1.75	2.12	2.58	2.92	4.01
17	0.257	0.534	0.863	1.33	1.74	2.11	2.57	2.90	3.97
18	0.257	0.534	0.862	1.33	1.73	2.10	2.55	2.88	3.92
19	0.257	0.533	0.861	1.33	1.73	2.09	2.54	2.86	3.88
20	0.257	0.533	0.860	1.33	1.72	2.09	2.53	2.85	3.85
21	0.257	0.532	0.859	1.32	1.72	2.08	2.52	2.83	3.82
22	0.256	0.532	0.858	1.32	1.72	2.07	2.51	2.82	3.79
23	0.256	0.532	0.858	1.32	1.71	2.07	2.50	2.81	3.77
24	0.256	0.531	0.857	1.32	1.71	2.06	2.49	2.80	3.75
25	0.256	0.531	0.856	1.32	1.71	2.06	2.49	2.79	3.73
26	0.256	0.531	0.856	1.31	1.71	2.06	2.48	2.78	3.71
27	0.256	0.531	0.855	1.31	1.70	2.05	2.47	2.77	3.69
28	0.256	0.530	0.855	1.31	1.70	2.05	2.47	2.76	3.67
29	0.256	0.530	0.854	1.31	1.70	2.05	2.46	2.76	3.66
30	0.256	0.530	0.854	1.31	1.70	2.04	2.46	2.75	3.65
40	0.255	0.529	0.851	1.30	1.68	2.02	2.42	2.70	3.55
50	0.255	0.528	0.849	1.30	1.68	2.01	2.40	2.68	3.50
100	0.254	0.526	0.845	1.29	1.66	1.98	2.36	2.63	3.39
> 100	0.253	0.525	0.842	1.28	1.65	1.96	2.33	2.58	3.30

Table 1: The quantiles of the t-distribution

Table A.6 Critical values of the F distribution ($\alpha = 0.05$; df_1 = treatment degrees of freedom, df_2 = error degrees of freedom). Part A: $\alpha = 0.05$

df_2 \ df_1	1	2	3	4	5	6	7	8	9	10	11	12
2	18.5	19.0	19.2	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.77	3.73	3.68	3.64	3.60	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.21	2.16
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.04
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83

Figure 2: The quantiles of the F-distribution