Final Exam

Introduction to Statistics(통계의 이해 영어강의) **2025 2nd semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

- All solutions must include a detailed step-by-step explanation.
- If an answer has more than four decimal places, round to the **fourth decimal place**.
- **Reference table** is provided on the last page of the exam.

1. Suppose we obtain a random sample $X_1, X_2, \ldots, X_n$ from a population with probability density function

$$f(x) = \begin{cases} \dfrac{2}{\theta}x, & 0 < x < \sqrt{\theta}, \quad (\theta > 0), \\ 0, & \text{otherwise}, \end{cases}$$

where θ is an unknown parameter. Answer the following questions. [10 points]

$\left(\text{Hint: } E(X_1^2) = \dfrac{\theta}{2},\ Var(X_1^2) = \dfrac{\theta^2}{12}.\right)$

(1) Find the constant k such that the estimator

$$\hat{\theta} = \frac{k}{n}\sum_{i=1}^{n} X_i^2$$

is an unbiased estimator of θ.

(2) Using the value of k from (1), show that $\hat{\theta}$ is a consistent estimator of θ.

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

2. A company manufactures bolts whose lengths (in centimeters) follow a normal distribution. A random sample of $n = 20$ bolts is selected, and the measured bolt lengths are shown in the table below. The sample mean and sample standard deviation are

$$\bar{x} = 16.29, \quad s = 0.084.$$

Construct a **90%** confidence interval for the population mean bolt length μ. [10 points]

Index i	1	2	3	$\cdots$	18	19	20
Bolt length x_i	16.25	16.30	16.25	$\cdots$	16.32	16.29	16.27

Final Exam

Introduction to Statistics(통계의 이해 영어강의) **2025 2nd semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

3. A trainer at Gym W claims that an exercise program **increases** muscle mass. For each of $n = 10$ participants, muscle mass (in kg) is measured **before** and **after** completing the program, and the difference is recorded. Let μ_{before} and μ_{after} denote the population mean muscle mass before and after the program, respectively, and define

$$\mu_{\text{diff}} = \mu_{\text{before}} - \mu_{\text{after}}.$$

A **paired t-test** is conducted at the significance level $\alpha = 0.05$. [15 points]

(1) State the null and alternative hypotheses. (Use a **one-sided** test.)

(2) Find the null distribution of the test statistic.

(3) The muscle mass measurements (in kg) before and after the program are shown in the table below. For the paired differences, the sample mean and sample standard deviation are $\bar{x}_{\text{diff}} = -1.8000$ and $s_{\text{diff}} = 2.9740$. Compute the observed test statistic.

Participant	1	2	3	4	5	6	7	8	9	10
Before	28	20	22	30	35	32	27	30	31	33
After	30	20	21	35	41	32	25	34	36	32
Diff. (Before − After)	-2	0	1	-5	-6	0	2	-4	-5	1

(4) Compute the p-value and complete the hypothesis test. State the conclusion in the context of the data.

Final Exam

Introduction to Statistics(통계의 이해 영어강의) **2025 2nd semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

4. A company wants to determine whether the mean **signal strength** (in dBm) measured from two antennas, A and B, is **different**. To do this, signal strength was measured $n_A = 22$ times using antenna A and $n_B = 26$ times using antenna B. Let μ_A and μ_B denote the population mean signal strength of antennas A and B, respectively. We conduct a **two-sample t-test** at the significance level $\alpha = 0.01$. [10 points]

(1) State the null and alternative hypotheses. (Use a **two-sided** test.)

(2) Find the test statistic and its null distribution.

(3) Using the following R code and output, report the p-value and complete the hypothesis test. State the conclusion in the context of the data.

```
Adat <- c(-98.70, -81.70, -83.00, # ... remaining values omitted ...
          -96.20)
Bdat <- c(-84.00, -76.21, -83.50, # ... remaining values omitted ...
          -79.10)
t.test(Adat, Bdat, alternative = "two.sided", var.equal = FALSE, mu = 0)
```

```
Welch Two Sample t-test

data: Adat and Bdat
t = -0.066942, df = 29.2, p-value = 0.9472
alternative hypothesis: true difference in means is not equal to 0

(...omitted...)
```

Final Exam

Introduction to Statistics(통계의 이해 영어강의) **2025 2nd semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

5. A food scientist wants to test whether **Brand 1 wine has a lower mean pH level** than Brand 2 wine. To test this at the significance level $\alpha = 0.05$, independent random samples of size $n_1 = 50$ and $n_2 = 50$ are selected from Brand 1 and Brand 2, respectively. Assume the pH levels follow

$$X_{11}, \cdots, X_{1,50} \sim N(\mu_1, 0.01), \qquad X_{21}, \cdots, X_{2,50} \sim N(\mu_2, 0.01),$$

where μ_1 and μ_2 denote the mean pH levels of Brand 1 and Brand 2, respectively. [15 points]

(1) State the null and alternative hypotheses. (Use a **one-sided** test.)

(2) Find the test statistic and its null distribution.

(3) Find the rejection region.

(4) Given $\bar{x}_1 = 3.4$ and $\bar{x}_2 = 3.5$, complete the hypothesis test, and state the conclusion in the context of the data..

(5) Find the power of the test assuming $\mu_1 - \mu_2 = -0.05$.

Final Exam

Introduction to Statistics(통계의 이해 영어강의) **2025 2nd semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

6. A used car dealer collected data on $n = 32$ used cars. For each car, the dealer recorded the car's **age** (x, in years) and the **selling price** (y, in dollars). We will use linear regression to predict selling price from age.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \cdots, n.$$

The summary statistics are as follows. [15 points]

$\bar{x}$	s_x	$\bar{y}$	s_y	r
6	2	14000	3000	-0.80

(1) Compute the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

(2) A used car is $x_i = 9$ years old and its actual selling price is $y_i = 10000$ dollars. Use the model to predict the mean selling price $\hat{y}_i$ at $x_i = 9$ and compute the residual e_i.

(3) Calculate the coefficient of determination R^2 and the residual standard error $\hat{\sigma}$.

Final Exam

Introduction to Statistics(통계의 이해 영어강의) **2025 2nd semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

7. Data on **12 students** consist of an intelligence test score (score1) and a chemistry score (score2). The chemistry score is treated as the response variable, y_i, and the intelligence test score as the predictor variable, x_i. We consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2).$$

Answer the following questions based on the following R output. [10 points]

```
score1 <- c(65, 50, 55, 65, 55, 70, 65, 70, 55, 70, 50, 55)
score2 <- c(85, 74, 76, 90, 85, 87, 94, 98, 81, 91, 76, 74)
fitted.model <- lm(score2 ~ score1); summary(fitted.model)
```

```
(...omitted...)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   30.0433    10.1366   2.964  0.014194 *
score1         0.8972     0.1665   5.389  0.000306 ***
---
Signif. codes: 0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

Residual standard error: 4.319 on 10 degrees of freedom
Multiple R-squared: 0.7438, Adjusted R-squared: 0.7182
F-statistic: 29.04 on 1 and 10 DF, p-value: 0.0003064
```

(1) Fill in the values shown in the table below.

$\hat{\beta}_0$	$\hat{\beta}_1$	**Estimated Regression Equation**	$MSE = \hat{\sigma}^2$

(2) Using the R output from part (1), test whether the intelligence test score (score1, x_i) is a statistically significant predictor variable for the chemistry score (score2, y_i) at the significance level $\alpha = 0.01$. Answer the following.

(a) State the null and alternative hypotheses.

(b) Report the p-value, and complete the hypothesis test. State the conclusion in the context of the data.

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

8. Consider the simple linear regression model

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

where $x_1, \cdots, x_n$ are assumed to be fixed constants. The least squares estimator of the slope β_1 is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Using the fact that

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0, \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})x_i,$$

show that $\hat{\beta}_1$ is an unbiased estimator of β_1. [10 points]

Reference Table

Satterthwaite's df : $\psi = \dfrac{(s_A^2/n_A + s_B^2/n_B)^2}{(s_A^2/n_A)^2/(n_A-1) + (s_B^2/n_B)^2/(n_B-1)}$, $S_p^2 = \dfrac{(n_A-1)S_A^2 + (n_B-1)S_B^2}{n_A+n_B-2}$

Critical Values for Standard Normal Distribution: z_α

α	**0.20**	**0.10**	**0.05**	**0.04**	**0.03**	**0.025**	**0.02**	**0.01**	**0.005**	**0.0005**
z_α	0.8416	1.2816	1.6449	1.7507	1.8808	1.9600	2.0537	2.3263	2.5758	3.2905

One-sided p-**values** from the standard normal distribution: $P(Z > z),\ Z \sim N(0,1)$

Base z	**Add** 0.00 **to** 0.09									
	0.00	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

Critical Values for t-Distribution: $t_\alpha(df)$

df	α									
	0.20	**0.10**	**0.05**	**0.04**	**0.03**	**0.025**	**0.02**	**0.01**	**0.005**	**0.0005**
17	0.8620	1.3334	1.7396	1.8620	2.0010	2.1098	2.2340	2.5669	2.8982	4.0259
18	0.8600	1.3304	1.7341	1.8567	1.9940	2.1009	2.2227	2.5524	2.8784	3.9651
19	0.8580	1.3277	1.7291	1.8516	1.9871	2.0930	2.2120	2.5395	2.8609	3.9095
20	0.8570	1.3253	1.7247	1.8471	1.9808	2.0860	2.2010	2.5280	2.8453	3.8590
21	0.8560	1.3232	1.7207	1.8430	1.9749	2.0796	2.1910	2.5180	2.8314	3.8120
22	0.8550	1.3212	1.7171	1.8393	1.9695	2.0739	2.1810	2.5083	2.8188	3.7680
23	0.8540	1.3195	1.7139	1.8360	1.9645	2.0687	2.1720	2.5000	2.8073	3.7260

One-sided p-**values** from the t distribution: $P(T > t),\ T \sim t(df)$

t	df											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1.80	0.161	0.107	0.085	0.073	0.066	0.061	0.057	0.055	0.053	0.051	0.050	0.049
1.82	0.160	0.105	0.083	0.071	0.064	0.059	0.056	0.053	0.051	0.049	0.048	0.047
1.84	0.158	0.104	0.082	0.070	0.063	0.058	0.054	0.052	0.049	0.048	0.046	0.045
1.86	0.157	0.102	0.080	0.068	0.061	0.056	0.053	0.050	0.048	0.046	0.045	0.044
1.88	0.156	0.100	0.078	0.067	0.059	0.055	0.051	0.048	0.046	0.045	0.043	0.042
1.90	0.154	0.099	0.077	0.065	0.058	0.053	0.050	0.047	0.045	0.043	0.042	0.041
1.92	0.153	0.097	0.075	0.064	0.056	0.052	0.048	0.046	0.044	0.042	0.041	0.039
1.94	0.151	0.096	0.074	0.062	0.055	0.050	0.047	0.044	0.042	0.041	0.039	0.038
1.96	0.150	0.095	0.072	0.061	0.054	0.049	0.045	0.043	0.041	0.039	0.038	0.037
1.98	0.149	0.093	0.071	0.059	0.052	0.048	0.044	0.042	0.040	0.038	0.037	0.036
2.00	0.148	0.092	0.070	0.058	0.051	0.046	0.043	0.040	0.038	0.037	0.035	0.034
2.02	0.146	0.090	0.068	0.057	0.050	0.045	0.042	0.039	0.037	0.035	0.034	0.033