

You have 120 minutes. There are eight questions. Each question is worth eight points. Books, notes, cheat sheets, calculators and cell phones are NOT allowed. Some tables and formulas are provided on the last pages of the exam.

Q 1 Twenty Koç students were asked what they did during the week before the final exams: stay in the dorms, go back to their family home, took a trip, etc. Their answers were as follows.

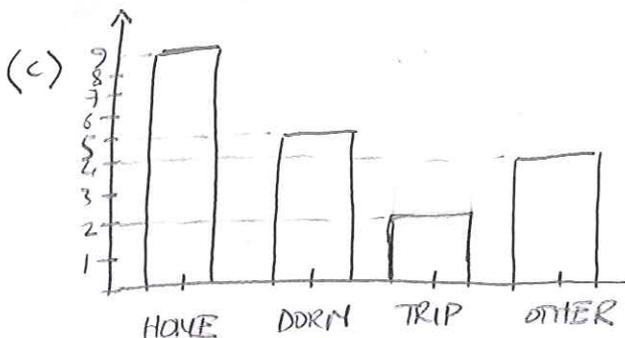
HOME	DORM	HOME	OTHER	DORM
DORM	HOME	TRIP	HOME	OTHER
OTHER	TRIP	HOME	HOME	DORM
DORM	HOME	HOME	OTHER	HOME

- (a) Is this data set qualitative or quantitative?
- (b) Compute the relative frequency of each answer.
- (c) Construct a bar graph and a pie chart for this data set.
- (d) Can we talk about the mean, the median and/or the mode of this data set?

(a) Qualitative.

(b)  $f(\text{HOME}) = \frac{9}{20}$  ;  $f(\text{DORM}) = \frac{5}{20} = \frac{1}{4}$  ;

$f(\text{TRIP}) = \frac{2}{20} = \frac{1}{10}$  ;  $f(\text{OTHER}) = \frac{4}{20} = \frac{1}{5}$  .



(d) Mean and median are not applicable for qualitative data  
 Mode is applicable; It is HOME.

Q 2 Ash, Berna and ten other Koç students take three taxis from Haciosman to Rumelifeneri.

- (a) What is the total number of ways that this group of twelve students can be partitioned to take three taxis? (Assume that each taxi accepts four passengers and we distinguish between the taxis.)
- (b) If this group is partitioned (to take there taxis) in a random way, what is the probability that Ash and Berna take the same taxi?

$$(a) \quad \# \text{ of partitions} = \frac{12!}{4! 4! 4!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{\cancel{4} \cdot \cancel{3} \cdot \cancel{2} \quad \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \quad \cancel{4} \cdot \cancel{3} \cdot \cancel{2}}$$
$$= 11 \cdot 5 \cdot 9 \cdot 7 \cdot 2 \cdot 5 = 34650.$$

(b) Ash sits in a taxi. There are a total of 11 seats left, 3 of which are in the same taxi with Ash. Therefore, the answer is  $\frac{3}{11}$ .

**Q 3** Bursaspor and Trabzonspor football (soccer) teams will have a match next week in the Turkish Super League. It is estimated that Bursaspor will win with 10% probability, Trabzonspor will win with 60% probability, and there will be a tie (beraberlik) with 30% probability. (Rule: A win brings three points to the winner and no points to the loser. A tie brings one point to each team.)

- (a) Calculate the expected number of points Trabzonspor will be awarded for this match.
- (b) Calculate the variance of the number of points Trabzonspor will be awarded for this match. (For your convenience,  $\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - \mu^2$ .)

(a) Let  $X = \#$  of points Trabzonspor is awarded.

$$\begin{aligned}\mu = E[X] &= \sum x p(x) = 0 \cdot (0.1) + 1 \cdot (0.3) + 3 \cdot (0.6) \\ &= 0.3 + 1.8 = 2.1.\end{aligned}$$

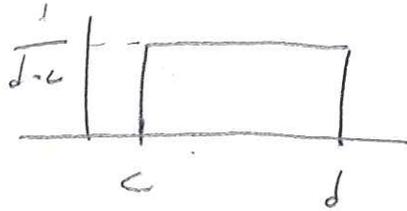
$$\begin{aligned}(b) \quad E[X^2] &= \sum x^2 p(x) = 0^2 \cdot (0.1) + 1^2 \cdot (0.3) + 3^2 \cdot (0.6) \\ &= 0.3 + 9(0.6) = 5.7.\end{aligned}$$

$$\begin{aligned}\text{Var}(X) &= E[X^2] - \mu^2 = 5.7 - (2.1)^2 \\ &= 5.7 - 4.41 \\ &= 1.29.\end{aligned}$$

Q 4 Let  $X$  be a (continuous) uniform random variable with mean 7 and variance 3.

(a) Write down the probability density function for  $X$ .

(b) Compute  $P(X \leq 6)$ .



$$a) \mu = \frac{c+d}{2} = 7; \quad \sigma^2 = \frac{(d-c)^2}{12} = 3.$$

$$c+d=14; \quad (d-c)^2=36; \quad d-c=6.$$

$$\begin{array}{l} d+c=14 \\ d-c=6 \end{array} \Rightarrow \begin{array}{l} c=4 \\ d=10 \end{array}$$

---

$$2d=20$$

$$f(x) = \begin{cases} \frac{1}{6} & \text{for } 4 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

$$(b) P(X \leq 6) = \frac{6-4}{10-4} = \frac{2}{6} = \frac{1}{3}.$$

Q 5 Let  $X_1, \dots, X_n$  be sampled independently from a common distribution with mean  $\mu$  and variance  $\sigma^2$ .

- (a) Is the sample mean  $\bar{X}$  an unbiased estimator of  $\mu$ ? Justify your answer.  
 (b) When  $n$  is sufficiently large, what is the approximate distribution of  $\bar{X}$ ? (Provide all of the parameters of this distribution.)  
 (c) When  $n = 2$ , simplify the sample variance  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ .  
 (d) When  $n = 2$ , is  $S^2$  an unbiased estimator of  $\sigma^2$ ? Justify your answer.

(a) YES.  $\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n} (X_1 + \dots + X_n)\right] = \frac{1}{n} E[X_1 + \dots + X_n] \\ &= \frac{1}{n} (\mu + \dots + \mu) = \frac{1}{n} n\mu = \mu. \end{aligned}$$

(b)  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ .

(c) 
$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2 \\ &= \left(\frac{X_1 - X_2}{2}\right)^2 + \left(\frac{X_1 - X_2}{2}\right)^2 = \frac{1}{2} (X_1 - X_2)^2. \end{aligned}$$

(d) YES. 
$$\begin{aligned} E[S^2] &= \frac{1}{2} E[(X_1 - X_2)^2] \\ &= \frac{1}{2} E[X_1^2 - 2X_1X_2 + X_2^2] \\ &= \frac{1}{2} (E[X_1^2] - 2E[X_1X_2] + E[X_2^2]) \\ \text{(by independence)} &\Rightarrow \frac{1}{2} (E[X_1^2] - 2E[X_1]E[X_2] + E[X_2^2]) \\ &= \frac{1}{2} \left[ (E[X_1^2] - \mu^2) + (E[X_2^2] - \mu^2) \right] \\ &= \frac{1}{2} (\sigma^2 + \sigma^2) = \sigma^2. \end{aligned}$$

**Q 6** Suppose there are 250 students taking the final exam. To estimate the mean score  $\mu$  of these students (out of 64 points) on this exam, the instructors sample 9 exams at random and then grade them.

- (a) If the sample mean is 34 and the sample variance is 25, write down the 90% confidence interval for  $\mu$ . Simplify your answer.
- (b) If the true variance is 24, at least how many exams should the instructors have graded so that the sampling error for the mean would be less than 1 point? (If you can simplify your answer and find the correct integer (tamsay), you will get a bonus point.)

(a) Sample size =  $n = 9$ . Use the  $t$ -table with  $df = 8$ .  
 (Assume that the score distribution of the whole class is approx. normal.)

$$\left[ \bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \right]$$

$$= [34 - 3.1, 34 + 3.1]$$

$$= [30.9, 37.1].$$

$$\alpha = 0.10, \alpha/2 = 0.05$$

$$t_{\alpha/2} = 1.86$$

$$s^2 = 25, s = 5$$

$$n = 9, \sqrt{n} = 3$$

$$t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1.86 \cdot \frac{5}{3}$$

$$= 3.1.$$

(b)  $SE = z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} \leq 1$

assuming  
 $n \geq 30$

$$n \geq (z_{0.05})^2 \cdot \sigma^2$$

$$= (1.645)^2 \cdot 24$$

$$= 64.9446. \quad \} \text{ Bonus}$$

Therefore, the answer is 65.  $\rightarrow$

Q 7 Istanbul Metropolitan Municipality states on its website that the duration of the Yenikapı-Haciosman metro ride is 27 minutes. However, you regularly commute from Yenikapı and suspect that the ride takes longer than 27 minutes on average. In order to test the official statement of the municipality, you take the metro 36 times (over the course of two months), record the duration of each ride, and then compute the sample mean and the sample variance.

- What are the null and alternative hypotheses in this example?
- What would constitute a Type I error in this example?
- If the level of significance of the test is supposed to be 0.10, what is the rejection region?
- If the sample mean is 27.5 and the sample variance is 4, what is the conclusion of the test?

(a) Null hypothesis  $H_0: \mu = 27$   
Alternative hypothesis  $H_a: \mu > 27$ .

(b) Type I error = rejecting the statement of the municipality when, in fact, it is true.

(c)  $n = 36$ , so we can use the z-statistic.  $\alpha = 0.10$

Rejection region:  $z > z_{\alpha} = 1.28$

$$(d) \quad z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{27.5 - 27}{2/\sqrt{6}} = \frac{1/2}{2/\sqrt{6}} = \frac{\sqrt{6}}{4} \\ = 1.5 > 1.28.$$

Conclusion: Reject  $H_0$  in favor of  $H_a$ .

**Q 8** You believe that the probability of a car being involved in an accident has nothing to do with its color. However, one of your friends claims that people who drive red cars are more prone to taking risks, whereas another friend argues that red cars are more visible by other drivers (and hence less likely to be hit). In order to test your hypothesis, you conduct an interview with 300 drivers, ask them the color of their car and whether they have been involved in an accident in the past five years. (Assume that drivers have not changed their cars during this period.) You find out the following:

- (i) 50 (of the 300 cars) are red, and 20 (of these 50 red cars) have been in at least one accident;
- (ii) 80 of the remaining 250 (non-red) cars have been in at least one accident.

Let  $p_1$  be the probability of a red car being involved in an accident in the past five years, and let  $p_2$  be the probability of non-red car being involved in an accident in the past five years.

- (a) Write down the point estimate for the target parameter  $p_1 - p_2$ .
- (b) Are the samples large enough to use the normal distribution (i.e., the  $z$ -statistic) for giving confidence intervals? Justify your answer.
- (c) Write down the 95% confidence interval for  $p_1 - p_2$ . (You do not have to simplify your answer.)
- (d) Is there sufficient evidence (at the 0.05 significance level) to conclude that the probability of being involved in an accident depends on the color of the car? Justify your answer.

$$(a) \quad \hat{p}_1 = \frac{20}{50} = 0.40; \quad \hat{p}_2 = \frac{80}{250} = 0.32; \quad \hat{p}_1 - \hat{p}_2 = 0.40 - 0.32 = 0.08.$$

$$(b) \quad \left. \begin{array}{l} n_1 p_1 = 20 \geq 15; \quad n_1(1-p_1) = 30 \geq 15; \\ n_2 p_2 = 80 \geq 15; \quad n_2(1-p_2) = 170 \geq 15. \end{array} \right\} \checkmark \text{ YES.}$$

$$(c) \quad (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$0.08 \pm 1.96 \sqrt{\frac{(0.4)(0.6)}{50} + \frac{(0.32)(0.68)}{250}}$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

(d) Next page.

Some statistics used for confidence intervals and hypothesis testing (under appropriate assumptions which you are supposed to know):

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - \mu}{s/\sqrt{n}};$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}};$$

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}};$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}};$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where} \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2};$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \approx \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}};$$

$$z \approx \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where} \quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

(d) Rejection region  $|z| > z_{0.025} = 1.96$

$$z \approx \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.08}{\sqrt{\frac{1}{3} \cdot \frac{2}{3} \left(\frac{1}{50} + \frac{1}{250}\right)}}$$

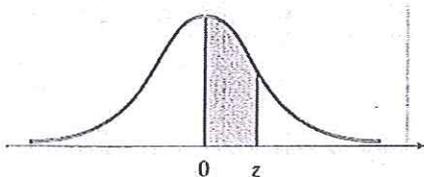
since  $\hat{p} = \frac{20+80}{50+250} = \frac{1}{3}$ .

$$= \frac{0.08}{\sqrt{\frac{12}{3^2 \cdot 250}}} = \frac{0.08 \cdot 5 \cdot \sqrt{10} \cdot \sqrt{3}}{\sqrt{2}} = 0.2 \cdot \sqrt{30} = \frac{\sqrt{30}}{5}$$

Therefore,  $|z| = \frac{\sqrt{30}}{5} \leq \frac{\sqrt{36}}{5} = \frac{6}{5} = 1.2 < 1.96$ .

There isn't sufficient evidence to reject the null hypothesis  $H_0: p_1 - p_2 = 0$ .

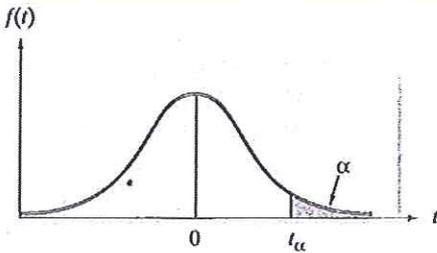
TABLE IV Normal Curve Areas



<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table I of A. Hald, *Statistical Tables and Formulas* (New York: Wiley), 1952. Reproduced by permission of A. Hald.

TABLE VI Critical Values of  $t$



Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
$\infty$	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Source: This table is reproduced with the kind permission of the Trustees of Biometrika from E. S. Pearson and H. O. Hartley (eds.), *The Biometrika Tables for Statisticians*, Vol. 1, 3d ed., Biometrika, 1966.