TERM TEST 1

Last Name:	**S**	**O**	**L**	**U**	**T**	**I**	**O**	**N**	**S**										
First Name:																			
Student #:																			

Instructor: M. Pivovarova

Duration: 90 minutes.

Allowed aids: A calculator and aid sheets provided (you may detach the aid sheets).

Enter your name and student number in the space provided at the top of the test paper.

Format: This test consists of 4 short-answer questions and one bonus question. Please, attempt the bonus question only after you are done with the main part of the tests. There are total of 105 possible points including the bonus question.

- For each question I give a guide for your response in brackets. It indicates what is expected: a quantitative analysis, a graph, and/or a written response. For example, "Is sampling error a plausible explanation for the result? [Analysis & 2 – 3 sentences]"
- Make sure to focus on answering the specific questions asked. Extraneous analysis does not earn positive marks even if it is correct and may earn negative marks if incorrect.
- Show your work and answer clearly, concisely, and completely. You do not have to fill all of the blank space: a generous amount is provided for your convenience.

	Q1	Q2	Q3	Q4	Bonus	Total	% Mark
Point Value	*25*	*25*	*25*	*25*	*5*	*105*	
Points Earned							

(1) [25 points] There are two large companies in Wonderland and you are considering to make an investment by buying their shares. One company, Monsters Inc, uses happiness and joy to produce sand castles. Another company, Troll and Co, recycles sadness to produce Halloween costumes. Below are the returns on shares for each of the companies in different times.

State of economy	Probability	Return on Investment (ROI)	
		Monsters Inc	Troll and Co
Merry Times	60%	15%	1%
Mediocre Times	30%	8%	6%
Really Bad Times	10%	-4%	16%

(a) [10 points] Find the expected return and risk for each company. Which company's return is more volatile relative to the mean return? (Hint: you should use a specific statistic to prove your opinion, but it is not a standard deviation or variance) [Analysis&1-2 sentences]

E[Monsters]=0.6*15+0.3*8+0.1*(-4)=11
V[Monsters]=0.6*(15-11)^2+0.3*(8-11)^2+0.1(-4-11)^2=34.8
SD[Monsters]=5.9

E[Troll and Co]=0.6*1+0.3*6+0.1*16=4
V[Troll and Co]=0.6*(1-4)^2+0.3*(6-4)^2+0.1(16-4)^2=21
SD[Monsters]=4.58

Note: both the variance and standard deviation are acceptable measures of riskiness.

CV (Coefficient of variation) measures the volatility of the return relative to the mean.
CV[Monsters]=5.6/11=0.51
CV[Troll and Co]=1.145

The return on Troll and Co is more volatile since CV[Troll]>CV[Monsters]

(b) [10 points] You finally made a decision and invested in 100 shares of which 60 are Monsters Inc and 40 are Troll and Co. If you know that the correlation between Monsters' and Troll's shares is -0.9, what is the expected return on your portfolio? What is the overall riskiness of your portfolio? [Analysis and 1 sentence]

E[Portfolio]=E[0.6*11+0.4*4]=8.2
V[Portfolio]=0.6^{2}34.8+0.4^{2}21+2*0.6*0.4*(-0.9)*5.9*4.58=4.21
SD[Portfolio]=2.05

(c) [5 points] A new company has been recently established in Wonderland. The good news is that the returns of that new company are robust to the state of the Wonderland economy, i.e. it has a constant 5% return. You consider selling you Troll and Co shares and instead buying the shares of the new company. What would be the riskiness and return of your new portfolio if the returns of Monsters Inc and the new company are not correlated?

E[New portfolio]=0.6*11+0.4*5=8.6
V[New portfolio]=$0.6^2$34.8+$0.4^2$0=12.528
SD[New portfolio]=3.54

(2) [25 points] Mr. Noxin is running for a dogcatcher. We know that in a population, 45% of all voters favour Mr Noxin.
(a) [10 points] Draw a table which describes the probability distribution of the number of people who favour Mr Noxin if 3 randomly selected individuals on a street are surveyed. Graph and carefully label the distribution. What is the probability that more than half will favor Noxin? [Your work, table and 1 number]

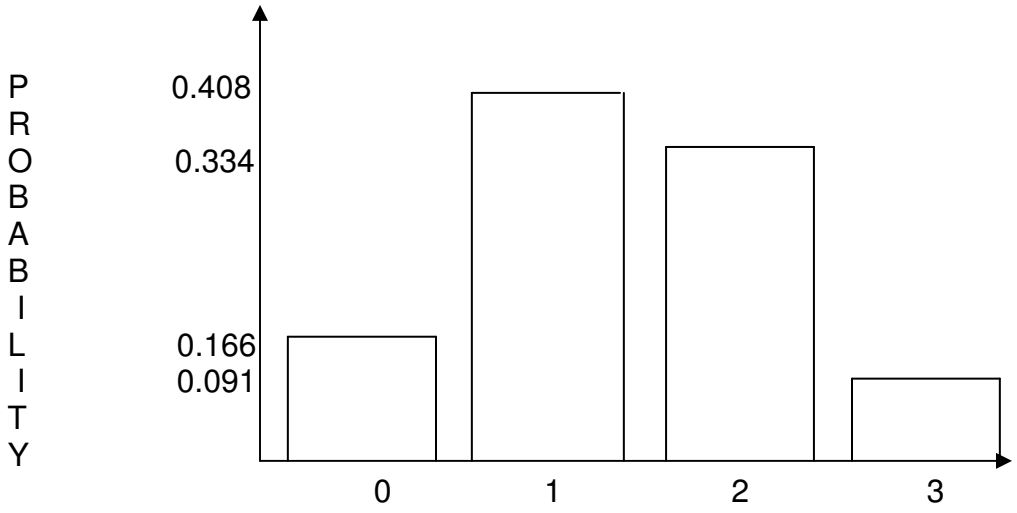
8 possible outcomes:

Outcome	FFF	FFN	FNF	NFF	FNN	NFN	NNF	NNN
P(outcome)	0.45^3	$0.45^{2*}0.55$	$0.45^{2*}0.55$	$0.45^{2*}0.55$	$0.45*0.55^2$	$0.45*0.55^2$	$0.45*0.55^2$	0.55^2
X	3	2	2	2	1	1	1	0

Probability distribution:

X	P(X=x)
0	0.55^3 = 0.166
1	$0.45^{*}0.55^2$ = 0.408
2	0.45^{2}*0.55 = 0.334
3	0.45^2 = 0.091

P(X>=2)=P(X=2)+P(X=3)= 0.334+0.091=0.425



(b) [10 points] There are 1,000,000 voters eligible to vote to replace the old dogcatcher. As before, only 45 percent of those eligible voters favour Noxin. But, because the weather is terrible and the race is boring, only 100 people bother to vote. If these 100 voters are a random selection from all eligible voters, what are Noxin's chances of winning? What if 1000 people vote? [Your work and 2 numbers]

If 100 people vote,

Find P(X>50)=?

First, check the rule of thumb:

np=100*0.45=45>10 and n(1-p)=100*0.55=55>10 >>> Can use normal approximation!

Mean of normal is np=45, standard deviation=sqrt(100*0.45*0.55)=4.975

P(X>50)=P(Z>[50-45]/4.975)=P(Z>1)=0.1587

If 1000 people vote,

P(X>500)=?

Mean is np=1000*0.45=450, st.dev.=sqrt(1000*0.45*0.55)=15.73

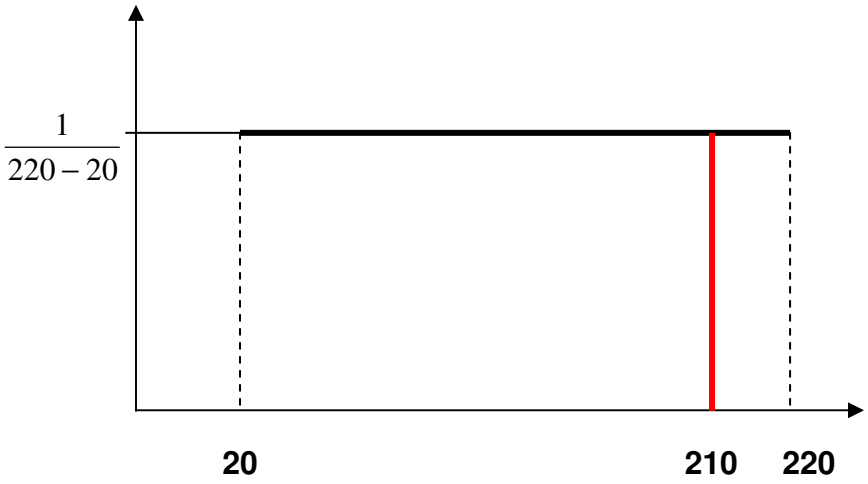
P(X>500)=P(Z>[500-450]/15.73)=P(Z>3.18)=0.0007

(c) [5 points] Can you provide any reasons for why the voter's turnout may not be an entirely random event, and thus may not satisfy the conditions for Binomial distribution? [3-4 sentences]

Decisions to vote might be correlated, i.e. not independent across individuals, thus violating one of the condition of the binomial distribution that trials should be independent.

(3) [25 points] You have been hired by a manager of a newly opened chain of supermarkets. The manager wants to improve customers' experience at checkout and proposes to give a 20% discount of the total bill amount if the checkout time exceeds a certain threshold. You have analyzed the data of the checkout times in this chain and found that the checkout time is uniformly distributed between 20 seconds and 3 minutes and 40 seconds.

(a) [7 points] If the manager wants to restrict the discount to no more than 5% of all customers, what time limit at checkout should he set? [Your work and 1 number]



$P(X>x_A)=0.05$

Probability is the area to the right of x_A. Area=Height*Width

Area=0.05, Height=0.005, Width=0.05/0.005=10

The cut-off point is 10 units away from the upper bound, 220-10=210 .

In order to have no more than 5% customers whose checkout time exceed a certain threshold, the time limit for the checkout time should be set to 3 mins 30 sec

(b) [6 points] How likely is that a randomly selected customer spends more than 3 minutes at checkout? What fraction of customers spends more than the average time at checkout? [Analysis and 2 numbers]

P(X>180)=Area=Height*Width=0.005(220-180)=0.005*40=0.2, or 20% chance that a customer will stay longer than 3 minutes

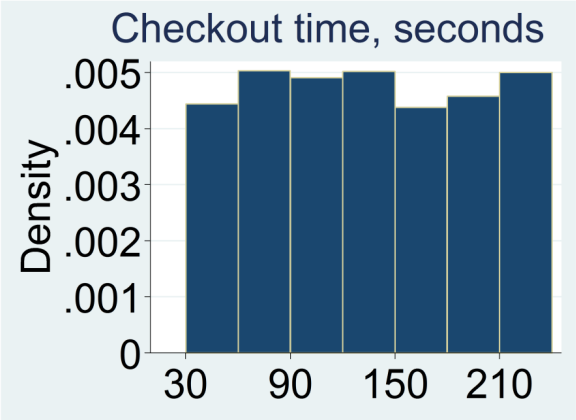
P(X>average)=P(X>[220+20]/2)=P(X>120)=Area=Height*Width=0.005*(220-120)=0.5, or 50%

Can answer immediately – the distribution is perfectly symmetric, the chance to be below or above the average is equal to 50%

►**For part (c)** Consider a histogram of the checkout times in one of the randomly selected supermarkets for a period of one week.

(c) [5 points] What fraction of customers in this supermarket will get a 20% discount according to the manager's proposal? [Analysis and 1 number]

We need to find P(time>210) which is the area of the bar to the right of 210. The height of the bar is 0.005 and the width of the bar is 30. The area is 0.005*30=0.15, or 15% of customers in this supermarket will get a discount if the manager's proposal is implemented.



(d) [7 points] Now consider that the total number of customers visiting this supermarket each week is 2000. You also know that the average bill in this supermarket is $32. What is the total expected value of the discount this supermarket will give to its customers per week if the manager's proposal is in effect? [Analysis and 1 number]

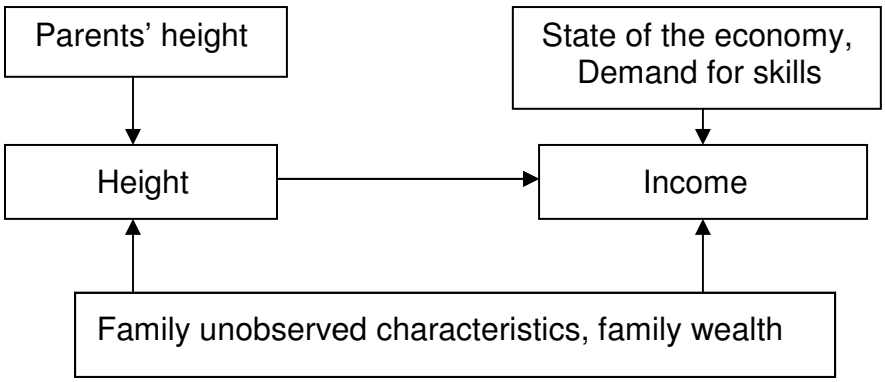
E[number of customers who get a discount]=0.15*2000=300

E[total amount spent by those 300 customers]=300*E[bill]=300*32=9600

E[discount]=E[Total amount*0.2]=0.2*E[total amont]=0.2*9600=1920

(4) [25 points] Consider the following abstract from an academic article: "There seems to be a societal impression that taller people are more successful in life …with income being the most common index of career success. … In this article, we conduct four investigations of the height-income linkage across the course of individuals' careers, using measures of height and income that are either longitudinal or independently reported."[1]

(a) [8 points] Please draw a flow chart similar to the one we saw in lectures which describes the research question in the article. Make sure to indicate the different types of variables and potential problems when analyzing this kind of relationship. [One flow chart and 2-3 sentences]



Here, height is an independent variable, and income is a dependent variable, family unobserved characteristics are lurking variables. We know that height is correlated with income, i.e. children from

[11] Journal of Applied Psychology 2004 (89) N 3, 428-441

poor families are more likely to be stunted because of the malnutrition in the early childhood. At the same time, children from poor families are more likely to drop out of school earlier, and have on average less years of education which translates into the lower income later in life.

(b) [7 points] The data were collected and a researcher estimated an OLS line for the relationship above in terms of z-scores as $\hat{z}_{Income} = 0.31 * z_{Height}$. You also know that the mean and standard deviation (in inches) of height variable are 69.09 and 3.27 respectively, and the mean and standard deviation of income variable (in thousands U.S. dollars) are 48.25 and 33.85, and n=151. Please write down an un-standardized OLS equation. Do not interpret the coefficients of your equation in this part! [Analysis and 1 equation]

$$b_1 = \frac{\text{cov}(Height, Income)}{\text{var}(Height)} = \frac{r(Height, Income) * sd(Height) * sd(Income)}{\text{var}(Height)} = \frac{r(Height, Income) * sd(Income)}{sd(Height)}$$

$$b_1 = \frac{0.31 * 33.85}{3.27} = 3.21$$

$$b_0 = Income\_bar - b_1 * Height\_bar = 48.25 - 3.21 * 69.09 = -173.53$$

Income_hut= -173.53+3.21*Height

(c) [5 points] Please interpret the coefficients of the OLS line in part (b). [3-4 sentences]

The intercept does not have any meaning in this equation – we do not have observations where height is equal to 0, so interpretation does not make any sense.

The slope of the OLS line indicates that income and height are positively related; on average, taller individuals earn more. We cannot claim that increase in height leads to increase in income, but we can say that on average individuals who are 1 inch taller have income which is 3.21 thousand dollars more.

(d) [5 points] What fraction of variation in income is explained by the variation in individuals' height? What is the SSE for this regression? [Your work and 2 numbers]

$R^2=r^2=0.31^2=0.0961$, or about 10% of the variation in income is explained by the variation in height.
To find SSE, remember that SST=SSE+SSR, and R^2=SSR/SST

Since $V(y) = \frac{\sum (y_i - \bar{y})^2}{n-1}$ **, then** $\sum (y_i - \bar{y})^2 = V(y) * (n-1) = 33.85^2 * 150 = 171873.375$

Since $R^2 = 0.0961 = \frac{SSR}{171873.375}$ **, then** $SSR = 0.0961 * 171873.375 = 1651.32$

SSE=SST-SSR=171873.375-1651.32=170222.055

(Bonus) [5 points] I offer you a gamble: you toss a fair coin 20 times and if you manage to get 15 or more heads in 20 tosses, I will pay you $900. In case the number of heads turns out to be less than 15, you pay me $10. What is the expected value of this game for you? [Your work and 1 number]

P(X>=15)=P(X=15)+P(X=16)+ P(X=17)+ P(X=18)+ P(X=19)+ P(X=20)=0.0207

E[game]=0.0207*900+(1-0.0207)*(-10)=8.837