Test II Review

Open Ended Questions

Problem 1

List the assumptions for the following hypothesis tests. Then write the equation defining the corresponding test statistic and degrees of freedom (if applicable):

- One-proportion z-test.
- Two-proportions z-test.
- χ^2-test for goodness of fit
- χ^2-test for independence
- Regression t-test
- Correlation t-test

Problem 2

Describe the the following concepts in the context of linear least squares regression. Don't just give formulas. Also give intuitive explanations

- Least squares criterion.
- Extrapolation.
- Correlation
- SST, SSR, SSE and r^2.
- The four conditions for performing inference with linear least squares regression.

Math Problems

For the following problems, indicate the degrees of freedom along with your test statistic and p-value (if applicable).

Problem 3

A coin was tossed and an eight-sided die (a "d8") was cast 80 times each. The coin and the die may be biased. The frequency tables are as follows:

	coin	
result	**Heads**	**Tails**
freq.	54	26

	d8							
result	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
freq.	9	8	12	8	15	6	13	9

a) Use a one-proportion z-test to test the hypothesis that the coin is biased at the 10% sig. level. Assume that the conditions for performing such a test are satisfied. (*Ans.* $z = \pm 3.13$, $p-$ value=0.001745, reject).

b) Use a χ^2 goodness-of-fit test to check the hypothesis that the die is biased at the 10% sig. level. Assume that the conditions for performing such a test are satisfied. (*Ans.* $\chi^2 = 6.4$ (df=7), $p-$value $= 0.4939$, fail to reject).

c) Redo part a) with a χ^2 goodness-of-fit test. (*Ans.* $\chi^2 = 9.8$ (df=1), same p-value. Note that $(\pm 3.13)^2 = 9.8$)

Problem 4

A study was conducting among PC gamers to determine if there is a relationship between the operating system (Linux, Windows) that they use and the brand of the GPU (AMD, Intel, Nvidia) that they use. The frequency table is as follows:

		Operating System	
		Linux	Windows
GPU Vendor	AMD	38	182
	Intel	34	108
	Nvidia	95	268

a) Conduct a χ^2 independence test to determine if there is any relationship between the type of operating system and the type of GPU used by PC gamers at the 10% significance level. Assume that the conditions for conducting such a test holds. (*Ans.* $\chi^2 = 6.20$ (df=2), p-value=0.04505, reject the null)

b) It is widely perceived that AMD GPU's traditionally perform worse on Linux than on Windows. Conduct a two-proportion z-test to determine if AMD GPU's are less popular with Linux users than with Windows users at the 10% significance level. Assume that the conditions for conducting such a test holds. (*Ans.* $z = \pm 2.43$, p-value=.0075, reject the null)

c) What problems might arise with testing multiple hypotheses with one dataset?

Problem 5

The following data contains paired observations of two numerical random variables x and y.

x	32	28	18	25	36
y	27	35	23	44	30

a) Calculate S_{xx}, S_{xy}, and S_{yy}. (*Ans.* $S_{xx}=118.8$, $S_{xy}=17.8$, $S_{yy}=262.8$)

b) Calculate the coefficients b_0 and b_1 for the linear least-squares regression $\hat{y}=b_0+b_1x$ as well as the corresponding value for r^2. (*Ans.* r^2=0.00639, $b_0=29.18$, $b_1=0.09428$)

c) Assume the conditions for performing inference for linear regression. Conduct a regression t-test to test the hypothesis that $\beta_1\neq 0$ against the null hypothesis that $\beta_1=0$ (*Ans.* $t=0.139$, p-value=.8983, fail to reject the nul).

d) Conduct a correlation t-test to test the hypothesis that $\rho\neq 0$ against the null hypothesis $\rho=0$ (*Ans.* $t=0.139$, $p-\text{value}=89.83\%$, fail to reject the null).

e) Can this linear regression be used to predict the value of y from the value of x?

Problem 6

Summary statistics for paired observations of two numerical random variables x and y are given below:

	x	y
sample size	40	40
mean	120	200
stdev	32	50
correlation	0.85	

(**Note:** you may have to use the formulas $S_{xx}=(n-1)s_x^2$ and $S_{yy}=(n-1)s_y^2$ as discussed in class)

a) Write the sample linear regression $\hat{y}=b_0+b_1x$ satisfying the least-squares criterion and the corresponding value for r^2. (*Ans.* $\hat{y}=40.63+1.328x$, $r^2=0.7225$)

b) Use the above regression to extrapolate a predicted value for y given the particular value for the predictor variable $x_P=105$. And provide a 90% prediction interval for this value. Assume that the conditions for performing such a hypothesis test are satisfied. (*Ans.* 180.08 ± 45.67)

c) What is the least-squares criterion?

d) In doing part b, you may have calculated that $s_e=26.68$. As discussed in class, this is an estimator of the population standard deviation of y conditioned on x. But this is much lower than the sample standard deviation of y, which is $s_y=50$. Why are they so different? Shouldn't they be the same?