EXERCISE ON OLS, PROBABILITY INTERVALS, ASYMPTOTICS

(1) For each of the $\{X_n\}$, Z pairs below, determine whether $X_n \xrightarrow[n\to\infty]{Q} Z$, where Q takes on the values P, D, $q.m.$, and $a.s.$. Explain your answers. For some of the cases where there is convergence in probability, determining whether there is a.s. convergence may be hard, so answers that do not succeed in determining a.s. convergence in such cases can receive full credit. You can use the central limit theorem and the strong law of large numbers:

CLT: If $\{X_n\}$ is i.i.d. with expectation zero and variance σ^2, $\sum_1^N X_n/\sqrt{N} \xrightarrow{D} N(0,\sigma^2)$.
SLLN: If $\{X_n\}$ is i.i.d with expectation μ, $\sum_1^N X_n/N \xrightarrow{a.s.} \mu$.

(a) $X_n = Z + Y_n$, where Y_n is i.i.d. $N(0,1/n)$

$$E[(X_n - Z)^2] = E[Y_n^2] = \frac{1}{n} \xrightarrow[n\to\infty]{} 0$$

Therefore we have $X_n \xrightarrow{q.m.} Z$. This implies convergence in probability and in distribution. This is one for which the $a.s.$ convergence question is somewhat painful, so you were not required to tackle it. For a.s. convergence, we must have that the probability that $|X_n - Z| < \varepsilon$ *and* $|X_{n+1} - Z| < \varepsilon$ *and* $|X_{n+2} - Z| < \varepsilon$ *and* . . . must go to one as $n \to \infty$. Since $X_n - Z = Y_n$, $P[|X_n - Z| < \varepsilon] = 1 - 2\Phi(-\varepsilon\sqrt{n})$, where Φ is the normal cdf. The probability that $|Y_{n+j}| < \varepsilon$ for all $j \geq 0$ is then

$$\prod_{j=0}^{\infty}(1 - 2\Phi(-\varepsilon\sqrt{n+j})) > 1 - \sum_{j=0}^{\infty} 2\Phi(-\varepsilon\sqrt{n+j})$$

Using l'Hôpital's rule, one can verify that $\Phi(\sqrt{n}\varepsilon)/e^{-.5\varepsilon n}$ converges to a constant as $n \to \infty$. But $\sum_{j=0}^{\infty} e^{-.5\varepsilon j} = 1/(1 - e^{-.5\varepsilon}) < \infty$. Since the whole sum is finite, the partial sum from n to ∞ has to go to zero with n. And because the ratio of $\Phi(-\sqrt{n}\varepsilon)$ to $e^{-.5\varepsilon n}$ goes to a constant, the sum from n to ∞ of $\Phi(-\sqrt{j}\varepsilon)$ eventually differs from the sum from n to ∞ by a factor of, say, between .9 and 1.1 (or $1 \pm \delta$ for any $\delta > 0$), $\sum_n^{\infty} \Phi(\sqrt{j}\varepsilon)$ is also bounded and goes to zero as $n \to \infty$. So we do have $a.s.$ convergence.

(b) X_n independent across n, $P[X_n = 1/n] = 1 - 1/n$, $P[X_n = n^2] = 1/n$, $P[Z = 0] = 1$.
Here $X_n - Z$ is just X_n. $E[(X_n - Z)^2] = 0 \cdot (1 - 1/n) + (1/n) * n^4 = n^3 \nrightarrow 0$. So we do not have $q.m.$ convergence. However, we have $P[X_n > \varepsilon] = 1/n$ for all $\varepsilon < 1$, and this does converge to zero. Therefore $X_n \xrightarrow{P} 0$, which in turn implies $X_n \xrightarrow{D} Z$, i.e. it converges in distribution to a discrete unit lump of probability at zero. Here again $a.s.$ convergence is hard. The probability that $X_j = 0$ for all $j > n$ is

$$\prod_{j=n}^{\infty}(1 - 1/j)\,.$$

But note that $1 - 1/n = (n-1)/n$, so $(1 - 1/n)(1 - 1/(n+1)) = (n-1)/(n+1)$. Successive numerators and denominators in the sequence cancel, so

$$\prod_{j=n}^{n+N} \frac{j}{j+1} = \prod_{j=n}^{n+N}(1 - 1/(j+1)) = \frac{n}{n+N}\,, \qquad \text{and therefore } \prod_{j=n}^{\infty}(1 - 1/(j+1)) = 0\,,\ \text{all } n\,.$$

Thus for every n, X_n exceeds 1 for some $j > n$ with probability one, and we do not have $a.s.$ convergence.

(c) X_n is the X_n from 1b, squared, $P[Z = 0] = 1$
Even though now both the mean and variance, and not just the variance, blow up to ∞ as n increases, the answers are exactly as before: No $q.m.$ convergence, no $a.s.$ convergence, but convergence in probability and in distribution.

(d) $X_n = \sum_1^n W_j/n + 1/n$, where $\{W_j\}$ is an i.i.d. sequence with $E[W_j] = \mu$, $P[Z = \mu] = 1$
The law of large numbers tells us that the time average of an i.i.d. sequence with finite mean converges $a.s.$ to its expectation, so $\sum_1^n W_j/n \xrightarrow{a.s.} \mu$, and $1/n$, being an ordinary sequence of numbers, converges $a.s.$ to zero. So their sum converges $a.s.$ to μ. Without knowing whether W_j has finite variance, you can't say whether it converges $q.m.$. But the $a.s.$ convergence implies convergence in probability and in distribution (to a degenerate distribution concentrated on the point μ.)

(e) $X_n = \sum_1^n W_j/\sqrt{n + \sqrt{n}}$, $\{W_j\}$ i.i.d. with mean zero and variance 1, $Z \sim N(0, 1)$.
The CLT tells us that

$$Z_n = \frac{\sum_1^n W_j}{\sqrt{n}} \xrightarrow{D} Z\,.$$

We can write

$$X_n = \sqrt{\frac{n}{n + \sqrt{n}}} Z_n\,.$$

But the factor in front of Z_n is a deterministic sequence that converges to one as $n \to \infty$, and thus also of course converges in probability to one. X_n is therefore a continuous function of one thing (the factor depending on n) that converges in probability to a constant and another thing that converges in distribution. One of the properties we listed in lecture was that such a function converges in distribution to the function evaluated at the limits. I.e. $X_n \xrightarrow{D} Z$. The problem gives no information about the joint distribution of the W_j's and Z, however, so no conclusion about convergence in $q.m.$, probability, or $a.s.$ is possible.

(2) Compute a linear regression of `testscr` on `str`, `avginc`, `meal_pct` `el_pct` `teachers` and `computer`, using the `caschool` data set.

(a) Form a χ-squared or F statistic to test the hypothesis that the coefficients on `teachers` and `computer` are both zero. Comment on whether it suggests these coefficients as a pair are unimportant in explaining `testscr`. [In R, the easiest way to do this is to get the package `car` (companion to applied regression), which you should be able to get with `install.packages(car)`. You then load the package with `library(car)`. After that several useful commands are available, including `linearHypothesis()`, which provides an easy way to generate test statistics like the one you are asked for here. With slightly more effort, you can do without `car` and use the function `vcov(lmout)` (to get the coefficient covariance matrix), then a matrix expression to get the chi-squared or F statistic, then `pchisq()` or `pf()`.]

(b) Find the estimated correlation matrix of the coefficients on these two variables, and use it to sketch a few representative level curves of their flat-prior joint posterior pdf in $\mathbb{R}^2$. [In R, `vcov(lmout)` delivers the coefficient covariance matrix for a linear regression, and `cov2cor(W)` converts the covariance matrix `W` to a correlation matrix.]

(c) Both `teachers` and `computer` are simple counts over the whole school, so they vary with the size of the school. If the size of the school were important, one might expect that the two variables would enter with the same sign. Another possibility is that it is computers *per teacher* that matters, so that we would expect the two coefficients to have opposite signs. Based on your sketch of the level curves, which type of effect would have higher posterior probability.