

STOCKHOLMS UNIVERSITET  
Statistiska institutionen  
Jörgen Säve-Söderbergh

**Skriftlig tentamen på momentet Undersökningsmetodik, 4.5  
högskolepoäng, ingående i kursen Regressionsanalys och  
undersökningsmetodik, 15 högskolepoäng, den 9 februari 2011, kl.  
9.00-14.00.**

**Skrivtid:** 5 timmar

**Hjälpmedel:** Godkänd miniräknare utan lagrade formler eller text. Vidhäftat  
formelblad och normalfördelningstabell.

**Tentamensgenomgång och återlämning:** Torsdagen den 3 mars, 14.00-15.00  
i B705. Därefter kan skrivningarna hämtas på studentexpeditionen, plan  
7 i B-huset.

Tentamen består av fem uppgifter som kan ge totalt 50 poäng. För full poäng  
krävs fullständiga och väl motiverade lösningar.

- 
1. En utländsk riskkapitalist söker en säker investering och funderar då på  
att starta ett taxibolag i en mindre svensk stad. Från stadens statistikkon-  
tor inhämtas uppgiften att det totala antalet taxiresor som görs i staden  
under ett år är 100 000. Under det följande året samlar riskkapitalisten  
in fullständiga uppgifter kring 400 taxiresor som görs under året. De 400  
färderna kan betraktas som ett obundet slumpmässigt urval utan återläg-  
gning från de 100 000 taxifärderna som görs totalt i staden.

Fördelningen över antalet passagerare per färd framgår av följande tabell:

Antal passagerare per färd	1	2	3	4	5
Antal taxifärder	145	180	50	20	5

- a) Beräkna ett 95%-igt konfidensintervall för det genomsnittliga antalet  
passagerare per taxifärd i staden under det aktuella året. (2 p)
- b) Beräkna ett 95%-igt konfidensintervall för det totala antalet passager-  
are som stadens taxibilar transporterade under året. (4 p)
- c) Beräkna ett 95%-igt konfidensintervall för det totala antalet färder  
under året där behov av stor bil har förelegat. Detta kan antas gälla  
då antalet passagerare är tre eller fler. (4 p)
2. I en viss bransch omfattande 5000 företag önskade man uppskatta de årliga  
reklamkostnaderna. Företagen grupperades i tre strata efter deras storlek.  
OSU utan återläggning användes för att hämta stickprov ur respektive  
strata. Följande resultat erhöles (i hundratusentals kronor) då ett stickprov  
på fyrahundra företag valdes ut:

Stratum	$N_i$	$n_i$	$\bar{x}_i$	$s_i$
1	4000	200	4	1.5
2	800	100	12	5
3	200	100	50	16

Beräkna ett 95%-igt konfidensintervall för den genomsnittliga årliga reklamkostnaden för de 5000 företagen. (10 p)

3. Priset på sockerbeter beror på sockerhalten. Att fastslå sockerhalten i ett stort parti sockerbeter är inte möjligt. I syfte att skatta denna parameter tar vi ett stickprov på tio sockerbeter. Med hjälp av en biokemist mäter vi vikten och sockerhalten bland de uttagna sockerbetorna. Därefter väger vi partiet sockerbeter och finner att det väger totalt 6000 kilo.

Sockerbeta	Sockerhalt (kg)	Vikt (kg)
1	0.2198	1.4
2	0.1716	1.2
3	0.1139	0.9
4	0.1452	1.1
5	0.1843	1.2
6	0.2281	1.6
7	0.1932	1.1
8	0.1382	0.9
9	0.1926	1.2
10	0.2311	1.4

- a) Skatta den totala sockerhalten i hela partiet av sockerbeter med hjälp av en kvotskattning. (7 p)
- b) Antag att vi får reda på att partiet består av 4800 sockerbeter, dvs  $N = 4800$ . Använd denna information för att skatta den totala sockerhalten i hela partiet sockerbeter. (3 p)
4. (Denna fråga är en fortsättning på fråga 2) En statistiker tog del av undersökningen i fråga 2 och undrade hur stickprovet hade allokerats. Det visade sig att det fanns finansiella medel för att göra om undersökningen.
- a) Från sin erfarenhet på detta område säger statistikern att standardavvikelseorna borde vara lika med  $\sigma_1 = 5, \sigma_2 = 6$  och  $\sigma_3 = 15$ , där  $\sigma_i$  står för standardavvikelsen i stratum nummer  $i$ . Beräkna hur de fyrahundra observationerna ska allokeras i detta fall. (4 p)
- b) En annan statistiker säger att det är rimligare att förutsätta att standardavvikelseorna är lika. Beräkna hur de fyrahundra observationerna ska allokeras i detta fall. (4 p)
- c) Vilken allokering förespråkar du? (2 p)
5. a) Hur många olika stickprov är det möjligt att välja bland när  $N = 12$  och vi ska använda systematiskt urval för att välja  $n = 3$ ? (2 p)

- b) Ge exempel på vad som skulle betraktas som ett *bekvämlighetsurval*. (3 p)
- c) Vilka förberedelser är det lämpligt att göra inför en undersökning och konstruktion av ett frågeformuläret? (3 p)
- d) Vad betyder förkortningen *PuL*? (2 p)